“Some People Require Subtitles”: The Effect of Subtitles on Viewer Comprehension in TV Series

Lim, Jie Er

"Some People Require Subtitles"[1]: The Effect of Subtitles on Viewer Comprehension in TV Series

Jie Er Lim

Department of Humanities, Leiden University

MA Linguistics: Translation

Supervisor: Dr. T. Reus

Second Reader: Dr. A.G. Dorst

Word count: 18663

July 1, 2022

[1] Van Tiem, V. (2014). *Love Like the Movies.* Pocket Star Books.

Abstract

This thesis aims to investigate whether the presence of subtitles aids viewer comprehension. Additionally, as a subquestion, this thesis also investigates whether a longer subtitle viewing time would lead to more comprehension. A total of 22 Dutch students watched a randomly assigned a subtitled or a non-subtitled clip from the medical drama *Chicago Med* while their eye movements were tracked. After the eye tracking process, the participants were post-tested on comprehension and detail questions. The results show that the participants with subtitles perform better on the post-test than their counterparts without subtitles on a statistically significant level. In specific, the subtitle group performs better than the non-subtitle group on detail questions. However, contrary to expectations, subtitle viewing time could not be found to be correlated to the performance of the participants. Taken together, the results from the post-test indicate that the Dual Coding Theory is in effect when viewing subtitled audiovisual material. This means that subtitles positively influence the viewer's ability to register, recall, and understand information and details.

Keywords: subtitling, viewer comprehension, eye tracking

Acknowledgement

To say that this thesis is my *magnum opus* would be a stretch. However, this thesis is something I am very proud of, and I would not have been able to create it without the help of various people.

First, I would like to thank Dr. S. Santos Ângelo Salgado Valdez and Dr. A.G. Dorst. The moment I decided that I wanted to work with an eye tracker, I had a feeling that this choice would be a great challenge because, at that time, I only had very little knowledge of eye tracking. It was with the help of the articles and books that Dr. Valdez forwarded to me that I increased my knowledge of eye tracking, which further increased my enthusiasm. My thesis topic was truly gaining shape after Dr. Dorst introduced me to one of her PhD students who was working with an eye tracker for his research. Although this was only a simple referral, I was able to meet with someone who had experience working with an eye tracker and who could help me formalize my idea.

Next, I would like to thank Sjoerd Lindenburg. When I first entered the LUCL lab, I had no idea where to start. However, you showed me how to navigate the programmes and work with the eye tracker; even offering me your template to set up my experiment to save me an abundance of work and time. Suddenly, working with the eye tracker became less scary and more feasible. Although I, eventually, did not use the eye tracker at the LUCL lab, you have shown me the difficulties and considerations that come with working with an eye tracker.

Of course, I would like to thank my supervisor Dr. T. Reus. Academic writing is not one of my strong points and it is something that I have been very insecure of. However, with your guidance and helpful feedback, I was able to produce a thesis that I am proud of. I also appreciate your advice and quick replies whenever I was stuck on something. It was very reassuring to know that wise words would be returning soon after hitting the send button. There were also stressful times when things did not go according to plan. Despite everything the universe decided to throw at me, I was able to pull through with your moral support.

The last person I would like to give my sincere thanks to is Ties Schalij. Without your help, my research would not have such a strong analysis. My knowledge of statistics is incredibly limited, and I did not even know where to begin. When you started saying word such as “binomial distribution” or “null hypothesis” my head would start spinning. Now, I know what those words are; I not only know the denotation, but I also understand *why* I had to opt for a binomial distribution for my research and *when* I could reject a null hypothesis. I will not forget the countless nights you spent scribbling on your whiteboard to explain every single detail to me; revising and repeating it every time I could not understand it. Although we had parted ways at the end of this thesis, you continued to support me. Proofreading and correcting mistakes in the analysis to ensure that, even if I was unsure about my thesis, I at least could be proud of my analysis. Thank you for helping me with my thesis and thank you for the past 4 years; they truly were the happiest years of my life.

Table of Contents

1. Introduction

‘Don’t watch so much television, it won’t do you good,’ is an often-heard phrase by young children, and admittedly, I have been the recipient of that phrase before. Although my eyesight has deteriorated, I would argue that watching television has benefits.

Watching television is namely a very accessible, informal, and suitable language learning input. Indeed, language learning can even be further enhanced by using subtitles while watching television. The Dual Coding Theory hypothesises that subtitles will positively impact language learning, as connecting the visual and verbal cues will make language more comprehensible and memorable (Paivio, 1986). On the other hand, the Cognitive Load Theory expects that subtitles will only be extraneous cognitive load for the working brain and thereby retard language learning (Sweller, 1988).

Taking inspiration from Kruger and Steyn’s 2014 research into the effect of subtitle reading on academic performance, this thesis aims to investigate whether the presence of subtitles has a positive effect on viewer comprehension. To do this, the following research question and subquestions have been set out:

RQ: Does the presence of subtitles help the viewer better understand the specialised language in *Chicago Med*?

Q1 Does the presence of subtitles result in a higher post-test score?

Q2. Does a higher rate of subtitle viewing lead to a higher post-test score?

Based on previous literature, I expect that the presence of subtitles will have a positive impact on comprehension. To be exact, I expect that the presence of subtitles will lead to a higher post-test score. Moreover, I believe that a higher rate of subtitle viewing will also lead to a higher score on the post-test score. The post-test is a test that contains questions about the video material to examine the participant’s comprehension.

To test this, participants are sourced from various universities. The participants are chosen based on a set of criteria: they must be Dutch and between 18 and 28 years old, have a

VWO diploma, and, preferably, not wear glasses. First, a pre-test questionnaire is used to gather information on the viewing habits of the participants. Then, the participants watch a trailer about the show to familiarize themselves with the characters, setting, and situation. This is done to prevent the participants from searching for information in the first minutes of the video. After viewing the trailer, the participants are randomly assigned to the video with subtitles or without. While the participants watch the video, their eye movements are tracked using an online eye tracker during the viewing of the video. Importantly, this research uses an eye tracker to confirm whether the viewers are looking at the subtitles and not relying on their own language skills. Finally, the participants need to fill in a post-test that consists of three general comprehension questions and three detail questions to check their comprehension of the video.

It is essential to find new tools that can aid language learning, because as Hao et al. (2021) have pointed out, there only is a limited amount of classroom time where language learning is stimulated. However, if we take a step back from the classrooms and look at the situation we have on hand, we can see another reason why language learning resources are necessary. Due to global conflicts, such as the war in Ukraine, people are forced to leave for other countries. This change requires the acquisition of a new language in order to accommodate to the new environment.

This thesis has been divided into five chapters. This first chapter was an introduction to the topic of this thesis. Chapter 2 contains the literature review that provides background information on the subjects that are central to this research. Chapter 3 explains the methodology that was used for this research. Chapter 4 presents the analysis of the results and chapter 5 presents an interpretation of the results in the context of the literature, the limitations, and further recommendations.

2. Literature

This thesis aims to investigate whether the presence of subtitles can help with the comprehension of a television show. To do so, this literature review is divided into 6 sections that together supply the necessary background information. Section 2.1 discusses the genealogy and definition of subtitling and audiovisual translation. Section 2.2 goes into language acquisition, its decline, and television as a language learning input. Section 2.3 presents critical theories that support or oppose the use of subtitles for language acquisition. Section 2.4 deliberates on the usage of eye-tracking in research and the benefits thereof. Section 2.5 reviews subtitling research that are relevant for this thesis and it presents the objectives of this research. Finally, section 2.6 provides a summary overview of the main points of each section.

2.1 Audiovisual translation and subtitling

Although the field of audiovisual translation is relatively new compared to the vast history of, say, poetry and literary translation, this form of translation might be the one that we come into contact with the most. For example, if you have watched the newest episode of Bridgerton or cried about the power of friendship in an anime, you have come into contact with audiovisual translation.

But what, exactly, are audiovisuals and what, in turn, is audiovisual translation? This section defines audiovisual translation and explain its history and how it became an essential part of films and series. Moreover, this section concentrates on the definition of subtitling and what the audiovisual nature of subtitles implies for the practice itself. The final section explains why subtitling and dubbing are so divided in Europe and what this division could imply for language learning using subtitles.

2.1.1 Audiovisuals and audiovisual translation

The word itself might already suggest what audiovisuals are. Audiovisuals are products that "are made to be both heard (*audio*) and seen (*visual*) simultaneously but they are primarily meant to be seen" (Chiaro, 2012, p.1). Although we might immediately think of the audiovisual products that we consume through screens, theatre plays, operas, and comic books are audiovisuals as well. The latter of the group might seem out of place as comic books are static and might be more suitable under the category of literature. However, Chiaro (2012) explains that the series of frames and the conventional language of comic books allow the story to unfold like a movie and therefore "they can be placed on the interface between print text and screen products" (p.2).

Considering this definition of audiovisuals, audiovisual translation (AVT) is the "transfer from one language to another of the verbal components contained in audiovisual works and products" (Chiaro, 2012, p.1). It emerged as a response to the emergence of film. During the era of silent films, AVT was not as necessary yet as only the intertitles of the film needed to be translated. The need for AVT became clear with the emergence of talking films in the 1920s (Remael, 2010). In order to export the films made in Hollywood, the films needed to be translated so that the new linguistic barrier could be overcome. As a result, dubbing and subtitling techniques were developed to solve this problem. Nowadays, there are more modalities that fall under AVT. These include subtitles, dubs, voice-overs, and closed captions (Remael, 2010).

2.1.2 Subtitling

Subtitling is one of the main modalities for screen translation (Chiaro, 2012). It can be defined 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 (letters, inserts, graffiti,

> inscriptions, placards, and the like), and the information that is contained on the soundtrack (songs, voices off). (Diaz-Cíntas and Remael, 2014, p. 8)

The difficulty of subtitling lies in the fact that "audiovisual materials are meant to be seen and heard simultaneously [which means that] their translation is different from translating print" (Chiaro, 2012, p.1). To synchronize the subtitles with the actors on screen, subtitles are confined to two lines only (somewhere between 35-39 characters for the Latin alphabet) (Karamitroglou, 1997) and appear only on screen for six seconds per two complete lines (Díaz-Cintas, 2012).

Due to these limitations, subtitles are (almost) never an exact translation of the audio. Thus, translators often have to resort to reduction as a translation strategy (Díaz-Cintas, 2012. Choosing what to reduce or omit is difficult for (especially beginning) translators, particularly for culture-specific items that often need a footnote to explain. Going too far in reduction sometimes leads to mistranslations, which, in turn, create controversy, as was the case in the "Squid Game controversy" (Squid Game subtitles 'change' meaning of Netflix show, 2021). In this controversy, the translators of the Netflix show *Squidgame* omitted numerous cultural references and items to accommodate the non-Korean audience. However, by doing so, the depth and significance of various scenes and characters were taken away as well. Thus, while it is necessary to conform to the fleeting nature of subtitles, the temporal, spatial, and linguistic constraints together create an obstacle for translators and sometimes lead to mistranslations. Even though translators oftentimes manage to capture the essence of the narrative in the translation, it is unfortunately the mistranslations that gain the spotlight.

2.1.3 The use of subtitles in Europe and the implications for this study

In Europe, both dubbing and subtitling are used to translate foreign films. Chiaro (2012) explains that traditionally, Western Europe has been divided into a subtitling block that included Scandinavian and Benelux countries, Greece and Portugal, while the so-called "FIGS" countries (France, Italy, Germany, and Spain) made up the dubbing block (p.2)

The choice between the two methods depends on “economic, ideological and pragmatic factors in the respective target countries” (Remael, 2010, p. 12). Economic motivations stemmed from the fact that subtitling was a cheaper alternative to dubbing. Thus, countries from smaller language areas, such as the Netherlands, prefer subtitling over dubbing because of its relatively lower cost (Ivarsson, 2009). On the other hand, ideological motivations stemmed from the American film industry’s influence on the country’s language, culture, and political regimes. As a response, each country would "adopt its own protection measures and/ or censorship mechanisms” (Nowell-Smith and Ricci, 1998, as cited in Pérrez-González, 2009).

Notably, some changes have been made over the years. The cost-effectiveness of subtitling has allowed the practice to slowly permeate in dubbing dominant countries (Chiaro, 2012), while children’s shows are nowadays more often dubbed in traditionally subtitling countries (Perego et al., 2015). Nevertheless, the preferred methods of translation have not changed much over the years, which led to the situation now in which “these marked preferences mean that each nation’s audience expects foreign film to be presented in the nationally dominant mode to which they have become accustomed.” (Mera, 1998, p. 74). These television viewing habits could have implications for language learning through subtitles. Viewers who are used to dubbing only, for example, might have difficulty splitting their attention between the visuals and the subtitles when they are presented with a subtitled video. This split attention falls in line with the Cognitive Load theory which will be further explained under section 2.3.

2.2 Language learning through subtitled television series

When I was younger, learning new languages was fairly easy. Now, I have unfortunately lost that gift. Going from multilingual, I had to say goodbye to my knowledge and transform into a byelingual. This raises the question as to how language learning works and why we suddenly stop “being good at language learning”.

This section explains how language learning works and why television and subtitles are suitable tools for language learning. Subsection 2.2.1 goes over language acquisition and its decline. Subsection 2.2.2 explains why television is a suitable language learning input. This is be done by using Nation's (2007) five language learning input conditions. Finally, subsection 2.2.3 links the theories and explains how implementing subtitles can enhance language learning. The goal of this section is to explain the decline of language learning and highlight the usefulness of television and subtitles accommodate that decline.

2.2.1 Language acquisition and its decline

Infants seem to learn languages at an incredible speed and with absurd ease as "young children learn their mother tongue rapidly and effortlessly, [they go] from babbling at 6 months of age to full sentences by the age of 3 years and follow the same developmental path regardless of culture" (Kuhl, 2004, p.831). Scientists have discovered that infants can quickly pick up language due to their perceptual abilities. Kuhl (2004) explains that "[infants] learn rapidly from exposure to language, in ways that are unique to humans, combining pattern detection and computational abilities … with special social skills" (p. 831). However, what makes infants so incredible at language learning is their ability to discriminate between phonetic units—an ability that adults lose (Kuhl, 2004).

The speed and ease with which infants learn languages decline as they grow up, and by the time they have reached adulthood, language learning has become a time-consuming and challenging process. The stabilisation of the phonetic units can explain this decline in ease. Kuhl (2004) explains that infants are sensitive to the various phonetic units. They can discriminate between the phonetic units because there is no distinct system yet that allocates the phonetic unit to a language. The lack of a stable system is also why bilingual children have no difficulty acquiring two languages.

> Bilingual children are mapping two distinct systems, with some portion of the input they hear devoted to each language. At an early age, neither language is statistically

> stable, and neither is therefore likely to interfere with the other" (Kuhl, 2004, p. 840).

Thus, stabilisation of the phonetic units means that infants become less sensitive to new phonetic units, and it signals the decline language of learning.

We now know that the most sensitive period for language learning is early childhood. Although we might not have been able to add one and two, our young brains were incredibly perceptive to unique phonetic units. Once those phonetic units start to stabilize, we become less perceptive of new phonetic units and our language learning skill declines. Knowing the benefits of learning languages at a young age, is there then a way children can be easily exposed to stimulating multilingual input? Moreover, is there a way we can make language learning easy again?

2.2.2 Television series as language learning input

Because language learning declines as we age, researchers have investigated alternative ways to make language learning easier; one of those methods is watching television. Although watching television may seem counter-productive, it is an informal and accessible way that provides learners with authentic and natural language. Moreover, "compared to other sources of comprehensible input such as reading, TV can provide a large amount of input in a short time." (Pujadas & Muñoz, 2020, p.552). In his language programme, Nation (2007) has outlined five conditions indicating whether an input is suitable for language learning (Rodgers, 2013). These five conditions have been summarised by Rodgers (2013) as follows:

1. The input needs to be processed in large quantities.

It is safe to say that watching television series is one of the most popular leisure activities. Its presence is evident in its impact on our vocabulary, with new words emerging, such as "binge watching" or "Netflix and chill". The numbers show that one of the most popular streaming platforms, Netflix, had "192,95 [sic] paid subscribers worldwide as of the second quarter of 2020" (Stoll, 2022). Netflix is so popular that "some U.S. adults have even admitted to watching

TV and movies in public restrooms" (Stoll, 2022). Rightfully, as Rodgers (2013) notes, "if language learners were to spend even a portion of their L1 viewing time on L2 television they would be processing a large amount of input" (p.2).

2. The input should be familiar to the learners.

Television shows suit this condition quite obviously. As noted in the previous condition, watching television is the most preferred leisure activity. Moreover, there is a wide variety of television shows and genres; learners can choose a television programme that they are familiar with and use it as a language learning channel. Rodgers also argues that viewers can build up familiarity with the television programme by watching multiple episodes.

3. The input should be stimulating and provide contextual cues.

Rodgers explains that "in order for this to occur the form of input must be rich in context cues, and there must be ways of building background knowledge" (p. 2). He points out that the imagery and dialogue provide context cues for the learners, and the episodic nature of television allows the learners to gain background knowledge on the series. I would like to add that the popularity of television shows also stimulates learners to converse and discuss "the latest episode", hereby engaging with new vocabulary.

4. Only a tiny part of the vocabulary presented through the input can be unknown.

Language learning is retarded if the language of the input is unclear. Rodgers writes that the lexical coverage ranges from 90%-99% for comprehension, while it narrows to 95%-98% for language learning. Although Rodgers notes that the relation between the percentage of known vocabulary and the effect on comprehension of television has not been studied, I believe that it should be safe to say that the language in television series is comprehensible for the average viewer.

5. The learners should be interested in the input and want to understand it.

Aside from the popularity of watching series, Rodgers' (2013) research findings also support this condition. His research found that

> the participants were generally positive about language learning through viewing English-language television. Most participants thought that learning from television was at least a *Pretty Good Use of Time* and *Pretty Enjoyable*, making these the two highest-rated aspects. The majority of the participants thought that viewing television had at least a *Somewhat Useful* effect on their overall English ability, was at least *Pretty Useful* for language learning in general, and had *Somewhat Improved* their listening skills. (p. 156)

Considering the findings in his research, it is safe to say that learners are interested in the input and the efficiency and enjoyability of this input stimulate the desire to understand the input.

Taking an overview of the properties of watching television shows, we can see that the popularity and accessibility make it a suitable input for language learning. In an enjoyable manner, learners are presented with new information that stimulates them to learn. We can already note bits of this theory being put in practice; nowadays, there are various shows for children (who as we know have an aptitude for language learning) that incorporate foreign languages into their episodes e.g., Spanish in the US children's show *Dora the Explorer* (Gifford, 2000-2019) or Mandarin in the US children's show *Ni hao Kai-Lan* (Harrington, 2008-2011). Interestingly, dialects can be acquired through television shows as well. Although it has not been confirmed by linguistic experts, various instances of the "Peppa effect" have been noted where children in the US started speaking with a British accent as a result of watching *Peppa Pig* (Yang, 2021).

At the end of section 2.2.1, I posed the question whether there is a way children can be easily exposed to stimulating multilingual input. These shows showcase that people are aware of the infant sensitivity period we mentioned in the previous section and put Nation's theory is to practice.

2.2.3 Subtitles as language learning enhancers

Even though television series are efficient English language learning input, it can be enhanced by implementing subtitles. In his research, Vanderplank (1988) uncovered that the use of subtitles helps language learners with spelling, comprehension of fast, authentic speech

and accents, and language retention. Talaván Zanón (2006) adds that using subtitles in language learning activities helps learners with pronunciation, developing recognition skills, understanding of English context-bound expression, and bridging the gap between reading and listening skills.

Furthermore, "when dealing with the potential usefulness of video input and subtitles in second language learning, it is necessary to bear in mind the major effects of visual associations on memory and the mnemonic power of imagery." (Danan, 1992, as cited in Talaván Zanón, 2006). Talaván Zanón explains that subtitled videos combine text, imagery, and sound and that this connection "encourages strong associations for retention and language use" (p. 43). In turn, the mnemonic aspect of this connection could serve a purpose in vocabulary learning or, considering the results from Vanderplank's (1988) research, learners could better remember particularities of the language they did not understand and ask questions about them.

If we take these benefits and apply them to the four basic language skills (reading, writing, listening, and speaking), it becomes clear that subtitles enhance language learning. The presence of subtitles improves the learner's speaking skills as it demonstrates how certain words are pronounced. Furthermore, it helps the learner's writing skills since it shows the spelling of words. Subtitles can also help during listening exercises as they offer clarity of fast speech or accents. Reading subtitles also proves to have mnemonic benefits as learners seem to be able to recall words or particularities easier with the help of subtitles.

2.3 Theories for and against the use of subtitles in language learning

The preceding section explained how language is acquired and why watching television combined with subtitles is a suitable language acquisition method. However, although research points towards the benefits of subtitles in foreign language learning, there are theories that oppose this practice. This section goes over the fundamental theories and principles that support and oppose the use of subtitles in language learning. Subsection 2.3.1 introduces the Dual

Coding Theory by Allan Paivio. This theory advocates for the benefits of subtitle use for language learning. Subsections 2.3.2 and 2.3.3 introduce theories that oppose the use of subtitles.

2.3.1 Dual Coding Theory

The key theory supporting the use of subtitles for comprehension and language learning is the dual coding theory (DCT), which was first introduced by Allan Paivio (1986). This theory hypothesises that cognition involves two different subsystems: one verbal system, logogens, that deals with language, and a non-verbal system, imagens, that deals with non-linguistic objects and events (Paivio, 2006, p. 3). Paivio (2006) explains that "the systems are assumed to be composed of internal representational units … that are activated when one recognises, manipulates, of [*sic*] just thinks about words or things" (p. 3). He points out that these two subsystems are "connected to sensory input and response output systems as well as to each other so that they can function independently or cooperatively" (p. 3). However, combining both channels would make the concrete language more memorable and comprehensible (Sadoski, 2001, p. 264), which aids language learning.

DCT can be applied to audiovisual translation and language learning as well. Mayer and Sims (1994) have adapted and modified Paivio's DCT and offered a three-step process to explain how visual and verbal cues can be integrated into the viewer's working memory during language learning. First, the verbal explanation is presented to the viewer. The viewer then constructs a mental representation of verbal information in the working memory. Mayer and Sims (1994) explain that "the cognitive process of going from an external to an internal representation of the verbal material is called *building a verbal representation connection* (or verbal encoding)" (p. 390). After this, the process is be repeated for the visual explanation that the viewer receives. That, in turn, is called "*building a visual representational connection* (or visual encoding)" (Mayer and Sims, 1994, p. 390). Finally, the working memory will construct

referential connections between the verbal and visual representations. Indeed, the stimuli processed so closely with each other create a contiguity effect (Mayer & Sims, 1994).

Mayers and Sims go on in their paper to investigate different types of viewers. If DCT prevails, future research could investigate the different types of viewers and what those viewers would imply for the subtitles, e.g. could translators include parts of the source text that are initially deemed secondary depending on the viewers? How would a subtitled video on heart conditions for medical students differ from a similar video for patients? Obtaining more information on the reception of the translation and the viewers could help translators work around the strict spatial and linguistic limits.

In sum, the Dual Coding Theory hypothesizes that cognition consists of two separate systems; one that processes words and one that processes non-linguistic objects and events. Whenever we view words or images, our working memory builds visual and verbal representations. The working memory in turn will construct referential connections between these two representations. By combining both systems, the words we have read and stored become more memorable. In practice, learners will be able to recall words and narratives easier with the help of subtitles.

2.3.2 Cognitive Load Theory

On the contrary, the cognitive load theory (CLT) argues against the advantages of subtitles for language learning and memory. John Sweller first introduced CLT in 1988. In short, CLT correlates the amount of information the working memory can hold. Sweller (1988) explains that during problem-solving, people use structures—also called schemas—that allow them to identify the category of the problem quickly and which solution is needed (p. 259). If the schemas are repeatedly applied, they can become automated. Once the schemas are automated, the working memory load will be reduced as the automation will automatically steer the behaviour without it needing to be processed again in the working memory (van

Merriënboer & Sweller, 2006). However, the cognitive effort that is needed to solve a problem, schema acquisition, and automation are processes that each require a considerable amount of working memory load. Since the working memory cannot hold much information, the instructional design should avoid adding cognitive load that does not contribute to problem-solving.

The working memory load can be affected by the load of the learning task itself—called the intrinsic cognitive load—or how the information is presented—called the extraneous load (van Merriënboer & Sweller, 2005). The situation sketched in 2.1.3 is an example of extraneous load. In that subsection, an example was given of viewers who are only used to dubbing. If dubbing viewers are presented with subtitled material, the working brain would be presented with a new manner of information transfer. This new material becomes an extraneous cognitive load for the working memory, which would retard language learning. Hence, several instructional design principles have looked at the instruction from the perspective of CLT and redesigned it to lessen the extraneous load. Three of those design principles are relevant for the research into subtitles in language learning, namely the split attention effect, the modality effect, and the redundancy effect.

Sweller (2005) explains that "the split attention effect occurs when attention must be split between multiple sources of visual information that are all essential for understanding" (p. 27). In principle, having two sources of visual information (imagery and subtitles) would increase the viewer's cognitive burden and, as a result, retard language learning. Van Merriënboer and Sweller (2006) suggest that the multiple sources of information be replaced by just a single source of information to lessen the extraneous cognitive load.

The modality effect is the second design principle that opposes subtitle use for language learning. This effect also occurs under conditions with multiple sources of information. According to the modality effect, as explained by Castro-Alonso and Sweller (2020),

> the working memory includes two systems with limited capacity: (a) the *phonological loop*, managing the processing of auditory information, and (b) the *visuospatial sketch*

> *pad*, dealing with visual and spatial information. The multicomponent model indicates that auditory and visuospatial information to a substantial degree tend to be processed separately in these limited systems. (pp. 75- 76).

Due to the separate systems, learning will be overloaded if it is solely focused on one system. Thus, having both subtitles and imagery will more likely overload the visuospatial sketchpad. Considering the modality effect, using a multimodal information source (both visual and auditory in one source) reduces the extraneous cognitive load (van Merriënboer & Sweller, 2006).

Finally, the redundancy effect differs slightly from the split attention and modality effects. The redundancy effect does not deal with multiple sources that contribute to the understanding and learning of the material. Instead,

> it deals with multiple sources of information in which one source is sufficient to allow understanding and learning while the other sources merely reiterate the information of the first source in a different form. (Sweller, 2005, p. 27)

Although the imagery and audio of the audiovisual material are sources that offer complementary information and not reiterated information, the subtitles are, in a sense, a reiteration of the audio. Thus, extraneous cognitive load can be reduced by eliminating redundant information (Sweller, 2005, p. 27).

Thus, in contrast the Dual Coding Theory, the Cognitive Load Theory hypothesises that the addition of subtitles would not aid language learning since the addition of subtitles would overload the limited amount of information that the working memory can hold. To balance the amount of cognitive load of the working memory, Sweller has constructed design principles that explain how instructions can decrease extraneous load. According to Sweller's design principles, the addition of subtitles would force the learner to unnecessarily split their attention between multiple sources of information, overload the visuospatial sketchpad, and add extraneous cognitive load.

In practice, the combination of visual material overloaded with visual cues and subtitles could be too much for the working brain to process. Instances such as a hospital scene with multiple things happening and various people talking at the same time at a rapid tempo force the learner to focus their attention to the visuals and audio. Subtitles could be a distraction in this case and retard understanding and memory.

2.3.3 Mayer's multimedia learning principles

In addition to the Cognitive Load Theory, Mayer's multimedia learning principles also hint towards the downsides of using subtitles. In his paper on multimedia learning, Mayer (2002) presents nine principles for the success of multimedia learning: multimedia, spatial contiguity, temporal contiguity, coherence, modality, redundancy, pre-training, signalling, and personalization. Each principle is supplemented with a theoretical rationale and empirical evidence from his research. In short, the design principles hypothesise that multimedia learning will emerge fruitful if the audiovisual material is limited to only imagery and audio that are presented simultaneously in a conversational style.

I would like to highlight Mayer's redundancy principle as it specifically concerns subtitles. According to the redundancy principle "students learn better from animation and narration than from animation, narration, and on-screen text" (Mayer, 2002, p.28). Mayer argues that on-screen text could overload the visual channel. Although his test results showed that test learners without on-screen text performed better than the learners with on-screen text, I would counter that his results are not completely reliable, as the participant pool only consisted of two test learners. A greater number of participants showing better test results without on-screen text would have been more convincing. Nevertheless, this is a principle that needs to be considered as it targets subtitles. If the results of this research indicate that participants without subtitles obtain a higher post-test score than participants with subtitles, it would further solidify Mayer's redundancy principle.

2.4 Eye-tracking in research

Section 2.3 explained the theories and principles for and against subtitles for language learning. This section focuses on eye-tracking and the application thereof. First, subsection 2.4.1 starts at a base level and explains why it is so valuable to track eye movements. Subsection 2.4.2 goes into more detail and explore the various fields in which eye-tracking can be applied. Finally, I focus on the research for this thesis and clarify what data eye-tracking has given on reading and subtitle reading in subsections 2.4.3 and 2.4.4, respectively.

2.4.1 Why is tracking eye movements valuable?

A special someone once told me that the eyes are the window to the soul, to which I retorted that the eyes "provide insight into cognitive processes such as language comprehension, memory, mental imagery and decision making" (Richardson & Spivey, 2004, p. 2). Although my answer was not the most romantic answer it is scientifically accurate. This was uncovered by studying saccades. Duchowsky (2017) explains that "[s]accades are rapid eye movements used in repositioning the fovea [a small spot responsible for accurate vision] to a new location in the visual environment" (p. 40). These saccades occur between 3-4 times per second, and these frantic movements are due to the overwhelming amount of visual information available (Richardson & Spivey, 2004). Because saccades are closely related to attentional mechanisms, they can give insights into the abovementioned cognitive processes.

Aside from eye movements, there is also much to learn from focus, spatial attention, and scene perception. Focussing on a specific location helps with processing the stimuli at that location (Posner, 1980, as cited in Richardson & Spivey, 2004). Moving a saccade to a focus point is intuitive, but spatial attention can be cast elsewhere while the eyes focus on one location (Richardson & Spivey, 2004). This spatial attention is called covert attention. Findlay & Gilchrist (2003) explain covert attention as paying attention without moving the eyes. An example to clarify this: a driver can still react in time to a child running after a ball despite being

focused on the road, or a gamer can detect enemies ambushing them from the side despite being focused on what is in front of them.

What happens, then, when we are viewing a scene? Early eye movement research has found that the viewer's fixation would focus on interesting or informative areas and neglect blank or uniform spaces (Buswell, 1936, as cited in Richardson & Spivey, 2004). More recent research has found that viewers reliably focused on only a few areas. Contemporary research has attempted to uncover what determines these areas of interest, and two components have been identified: the statistical properties of an image (e.g., extremes or luminosity) and local contrast. These discoveries confirm the finding that viewers focus on interesting and informative areas rather than blank or uniform spaces. If this idea is applied to subtitling, we can already hypothesise that subtitles would attract the viewer's fixation as well; the contrasting colours of the letters, fleeting nature of subtitles, and translation create an interesting and informative area that attracts the viewer's attention. Indeed, this idea will be further expanded on subsection 2.4.4.

2.4.2 Areas of application

Because eye movement can indicate so much, eye movement research has proven to be fruitful in a variety of fields; they vary from commercial research into advertising and design (Scott et al., 2016; Ranney et al., 2019; Hervet et al., 2011) to academic research into neuroscience and psychiatry (Hessels & Hooge, 2019; Hannula et al., 2010). The motivation to use an eye tracker to investigate consumers' behaviour is relatively straightforward. By investigating the areas of interest and gaze duration, companies can deduce what consumers find exciting or what catches their attention to improve their advertisements. Companies can adjust the design of their website or advertisement to subconsciously guide the consumers through the website or catch their attention by implementing triggers, such as pop-ups. Researchers can also investigate whether the advertising is successful for non-commercial

purposes. For example, Ranney et al. (2019) used eye-tracking to investigate whether the anti-smoking advertisements on toxic chemicals in cigarettes were effective. They found that participants spent more dwelling time on advertisements containing text, which indicated that they spent significant time attending to the information and processing it. The participants chose the advertisements containing warnings about the toxic chemicals to be the most effective.

In neuroscience, eye movements have shown to be useful for different applications. Scientists have uncovered that eye movements can give insight into abnormal brain functioning. Richardson and Spivey (2004) present the findings by Diefendorf and Dodge (1908), who had observed that patients with "praecox dementia", now known as schizophrenia, had erratic eye movements. While it is true that the human eye moves frantically during saccades, the human eye can also slowly and smoothly move while following a moving object. Patients with schizophrenia proved not to be able to do so and had more saccade movements than the control subjects. Another interesting finding was that people with schizophrenia displayed different eye scanning movements and scan paths than the control group (Shimizu et al.,2000; Williams, Loughland, Green, Harris, & Gordon, 2003, as cited in Richardson & Spivey, 2004). In practice, collecting this information can help clinicians diagnose schizophrenia in children and help them before the onset of psychosis. These distinct areas of application showcase the versatility of eye tracking and the information it could provide.

2.4.3 Reading habits

Eye-tracking has provided plenty of information on how we read. We learn that while we read, eye fixations last approximately 200-250 milliseconds (Pollatsek, Rayner, & Collins, 1984, as cited in Richardson & Spivey, 2004). The reading saccades cover seven to nine letters depending on the font size and distance (Morrison & Rayner, 1981, as cited in Richardson & Spivey, 2004). If we look at shorter words, consisting of 2-3 letters, they are skipped 75% of the time, while eight-letter words are almost always fixated. The chances of an individual

fixation also depend on the class of the word; content words are fixated 85% of the time, while a function word is only fixated 35% of the time (Carpenter & Just, 1983, as cited by Richardson & Spivey, 2004).

Eye-tracking has also given us insight into how we focus on words. Because the eye has such a small focus point, the text that falls outside the window becomes a blur. Interestingly, "[r]eading is unaffected, however, when the window extends 3-4 letters to the left of fixation and 14-15 letters to the right, suggesting that there is a perceptual span of around 18 characters centred asymmetrically around the fixation point" (Richardson & Spivey, 2004, p. 13). It has been uncovered that these asymmetrical centres are language-specific; the asymmetry is mirrored for languages that are read left to right and rotated for languages that are read vertically (Richardson & Spivey, 2004).

All of this data can be used to surmise the cognitive processing of a text. Tracking eye movements can reveal significant differences between individuals; different eye patterns can be discerned between children, people with dyslexia, and even graduate students or professors. These differences, in turn, could have considerable practical in educational psychology (McKane, Maples, Sellars, & McNeil, 2001, as cited in Richardson & Spivey, 2004). Nevertheless, this data is what we have obtained from participants reading static written text that can be re-read at one's own tempo. The temporal and short nature of subtitles does not allow for this, which influences the way we read on-screen text.

2.4.4 Eye-tracking and subtitle reading

Due to the fleeting nature of subtitles and the fact that viewers have to divide their attention between the moving subtitles and the moving visuals, researchers have found that the pattern of the viewer's eye movements has changed. Instead of focusing on the text only, Kruger et al. (2015) found that the presence of subtitles splits the attention from the viewers, causing them to flicker between the subtitles and the imagery. Although the scan path might show that

viewers pay attention to both imagery and subtitles, Jensema et al. (2000) found that reading the subtitles appeared to be a priority to the viewers. They discovered that whenever subtitles appeared on screen, the viewers changed their viewing patterns and moved their eyes to the middle of the screen before starting to read the subtitles and only returning to the visuals once they had finished reading. This shows that the eyes are drawn to subtitles because they identify subtitles as a source of information and because their changing nature triggers the eyes to flicker towards them (Kruger et al., 2015). The notion of the visual attraction of subtitles is further confirmed by Szarkowska et al. (2016) who found that "although many hearing participants were proficient in English and all of them were native speakers of Polish, they were still gazing a lot at both types of subtitles: they looked at as much as 83% subtitles in English videos and 76% in Polish videos" (p.200).

Although common ground has been established on subtitle reading, more research is needed. Simple assumptions such as "people are less likely to read intralingual subtitles because they already understand the language" cannot be relied on. The findings of Kruger et al's (2014) research showed that intralingual subtitles were less skipped than interlingual subtitles. In their research, Sesotho students were split into two groups. One group watched a recorded lecture in English with English intralingual subtitles, while the other group watched the same lecture but with Sesotho interlingual subtitles. The native language of the students was Sesotho. The results showed that the interlingual subtitles in Sesotho were skipped 50% on average, while the intralingual subtitles in English were only skipped around 20%. Meanwhile, the research by Szarkowska et al. (2016) showed reverse results. The research by Szarkowska et al. (2016) showed that the Polish participants looked less at the intralingual subtitles in Polish videos (4.05 fixations per subtitle, 851 ms spent in the subtitle area) and more at the interlingual subtitles in English videos (4.35 fixations per subtitle, 908 ms in the subtitle area) (p. 200). If there is anything that these two researches indicate, is that subtitle viewing is not as straightforward is

it might seem, and more research is needed on the processing of different language pairs and viewers.

Eye-tracking has given us an abundance of information on people's viewing habits and reading habits. By tracking saccades, we can uncover the cognitive processes that are behind the processing of the visual world around us. Moreover, we have uncovered that irregular eye movements could be an indication of abnormal brain functioning. Although people follow a certain pattern when they are viewing a scene or reading a text, "the addition of captions to a video resulted in major changes in eye movement patterns, with the viewing process becoming primarily a reading process" (Jensema et al., 2000, p. 275).

In the introduction of this subsection, I shared an anecdote about the time someone told me that the eyes are the window to the soul. Knowing the abundance of information eye tracking can give on our cognitive processes, I would like to say that eye tracking is the window to our soul.

2.5 Relevant subtitling research and study objective

The usefulness of captions is best demonstrated in the research by Markham (1989) on the effects of subtitles on listening comprehension of beginning, intermediate, and advanced ESL (English as a second language learner) students. Rationally, Markham had several expectations before the start of his research. He expected that captions would be less significant for his advanced participants compared to the other participants and that the beginning ESL students would have difficulty understanding the captions as they were still novice learners. Nevertheless, his advanced participants showed to have benefited as much from the subtitles as the other participants. Moreover, the beginning learners performed at a higher level when they were provided with subtitles.

Taking the research into the relation between captions and comprehension a step further, Kruger and Steyn (2014) integrated the use of an eye tracker into their research on the impact

of subtitles on comprehension. By measuring fixations and saccades and comparing those to the results of the post-test, Kruger and Steyn could not find a significant difference between the subtitle group and the non-subtitle group. However, because Kruger and Steyn separated the subtitle group between participants who read the subtitles and participants who did not read the subtitles fully, they found that the participants that did fully read the subtitles performed better than the participants that did not fully read the subtitles. Although Kruger and Steyn's research was not conclusive, it did hint towards the effectiveness of reading subtitles for comprehension.

The current research takes inspiration from Kruger and Steyn's research. It aims to investigate whether the presence of subtitles has a positive effect on the comprehension of viewers. Moreover, it aims to investigate whether a longer viewing time has a positive effect on the comprehension as well. This second inquiry is, however, different from Kruger and Steyn's research, as they measured the level of reading whereas this research focuses on viewing time. Although subtitle reading and subtitle viewing time are two different notions, I believe that Kruger and Steyn's findings can still be used to compare the findings of this research with. Kruger and Steyn discovered that the participants that read the subtitles fully performed better than the participants that did not read the subtitles fully. To read the subtitles fully, the participants would need to spend a longer time looking at the subtitles. Based on this, it can be assumed that a longer viewing time leads to a higher comprehension.

2.6 Summary of the literature

This literature chapter provides a background on various topics related to the current research into the effectiveness of subtitles in listening comprehension. Research into language acquisition has uncovered that during early childhood, our brains are the most perceptive to the phonetic units which makes language learning the easiest. Once those phonetic units have stabilized, the sensitive period closes, and language learning becomes more difficult.

To accommodate to the challenges of language learning and to work efficiently with the sensitive period, various forms of language learning input were investigated, one of which was watching television. Although it is a simple pastime, Nation's (2007) five conditions substantiate the argument that it a useful language learning input, as it is popular, familiar, stimulating, not too difficult to understand, and interesting to the learners. By adding subtitles to the audiovisuals, Vanderplank (1988) has demonstrated that language learning becomes easier, and language becomes more memorable. This falls in line with the core of Dual Coding Theory, which theorizes that combining verbal and non-verbal systems makes language more memorable and comprehensible. On the other hand, Cognitive Load Theory theorizes that the addition of subtitles will form extraneous load for the working memory which in turn retard language learning.

Despite the Cognitive Load Theory, research has indicated that subtitles do help with listening comprehension. Markham's (1989) research has indicated that despite the proficiency level of ESL students, subtitles still form substantial help for listening comprehension. The research into captions and comprehension was advanced by the addition of an eye tracker by Kruger and Steyn (2014). By employing the eye tracker, Kruger and Steyn were able to track the eye movements of the participants and investigate the fixation and saccades which showed what the participants were focussing on while watching the video. Although their research did not deliver decisive results, it did indicate that the participants who fully read the captions scored better than the participants who did not fully read the captions. Truly, this finding showcases the usefulness of eye-tracking; although all participants of the subtitle group viewed the subtitles, Kruger and Steyn were able to discern which participants fully read the subtitles and whether this would have implications for their research.

The theories that were discussed in this chapter form a fundamental base for this research into the relation between comprehension and subtitle reading. Nation's theory and Vanderplank's research substantiate why subtitled videos are a suitable language learning input.

Their usefulness is further corroborated by the Dual Coding Theory. This theory hypothesises that by combining the verbal system and the visual system in our working brain, language becomes more comprehensible and memorable. In case of adverse findings, the findings can be justified by the Cognitive Load Theory. According to this theory, overloading the working memory with extraneous load will retard language learning. In the following chapter, I discuss the methodology of this research. Since this research is a modification of Kruger and Steyn's research on a smaller scale, some alterations have been made. Those changes are explained and justified in the following chapter as well.

3. Methodology

This chapter discusses the method, participants, materials, and procedure of this study. The method for this research is explained in section 3.1. It details and justifies the changes that have been made to Kruger and Steyn's (2014) method. Section 3.2 provides an overview of the participants and explains the selection criteria. Section 3.3 explains the design of the pre-test. Section 3.4 discusses the materials and procedure. This includes the pilot test, videos, the eye tracking software, and the procedure. Finally, section 3.5 discusses the post-test and the binomial distribution, which is used to analyse the statistical significance of the results.

3.1 The adapted method from Kruger and Steyn

In their 2014 research, Kruger and Steyn investigated whether subtitle reading had a positive effect on academic performance. They did this by creating a reading index to uncover how much their participants read the subtitles. The current research was loosely based on Kruger and Steyn's (2014) research and adopted a variation of Kruger and Steyn's method. Several changes were made to suit a smaller-scale research and to adjust to the limited time frame. Instead of using an eye tracker at a lab, an online eye tracker was employed for reasons of practicality. Moreover, while Kruger and Steyn ran their experiment with ESL students, this study uses university students of various backgrounds to simplify and expedite the participant sourcing. Most importantly, as was mentioned in the Literature chapter, the analysis of this research focused on subtitle viewing rather than subtitle reading. This was done, as it was not possible to use Kruger and Steyn's Reading Index for Dynamic Texts (RIDT). Kruger and Steyn designed the RIDT "with the potential to provide a reliable measure of the visual processing of subtitles, or reading behavior of viewers over extended text" (2014, p.110). However, only short fragments are checked for this research rather than an entire video. Using the RIDT would thus yield unrepresentative results. Instead, the analysis for this research used the raw percentage of

observations in the AOI during the chosen time frame to investigate whether a higher subtitle viewing time leads to a higher post-test score.

3.2 Participants

This study included 23 participants. They ranged in age from 18 to 26, with an average age of 22. In total, 10 participants received the video without subtitles, and 12 received the video with subtitles. One participant was removed from the study because it appeared that they did not meet the criteria that are set out in the subsection below. The ineligibility of this participant was only discovered after the experiment had finished. Multiple participants were tested at the same time and due to this, the forms were only checked after the experiment was finished.

3.2.1 Selection criteria

The participants were selected on the basis of four different criteria:

- The participant had to be a native speaker of Dutch and a non-native speaker of English.

This criterium was chosen because the aim of this research is to investigate the functionality of subtitles in video comprehension of Dutch viewers. Moreover, it was important that the participant was a non-native speaker of English to ensure that the video material would not be too easy for them to understand.

- The participant had to have obtained at least a Dutch VWO high-school diploma.

Due to time constraints, it was not possible to create an English level assessment test for the participants. Instead, the participants were asked whether they had obtained a VWO high-school diploma and their current study programme. Participants who did not have a VWO diploma but did complete a bachelor's taught in English or completed a propaedeutic year at HBO (University of Applied Sciences) were also eligible.

All high-school national final exams in the Netherlands are regulated by the Tests and Examinations Board (College voor Toetsen en Examens) (Rijksoverheid, n.d.). According to this board, students will have B2 level listening skills at VWO level. This level means that students who have obtained B2 skills "can understand extended speech … [and they can] understand majority of films in standard dialect" (Council of Europe, 2001, p. 27). Based on this description, it was safe to assume that the participants with a VWO diploma would be able to understand a television series.

- The age range of the participant had to be between 18 and 28 years.

The decision was made to limit the age range to focus on a specific part of the population. This decision would ensure that the results could represent that age group. Furthermore, this age range was chosen because the researcher mostly contacted peers or other people from that age range, which made the participant recruitment process easier and more successful.

- It was desirable if the participant did not wear glasses.

The reflection and shine of the glasses could interfere with the retina detection of the eye-tracking software. In cases where the prescription of the glasses was low enough for the participant to read the subtitles without glasses, they could still participate. In that case, it was imperative that the participant did not squint their eyes, as this could interfere with the retina detection.

3.3 The pre-test questionnaire

The pre-test questionnaire was designed to check the participants' eligibility and to obtain data on the participants' viewing habits. The first questions were questions about the participants' age and education level. The second set included questions about media consumption. The participants were asked to indicate the amount of time they spent watching series, on which medium and streaming service they watched series, whether they used

subtitles, and what language the audio and subtitles were. Data on these topics could be used to substantiate later findings. The preliminary questionnaire can be found in Appendix A.

3.4 Materials and procedure

This subsection discusses the equipment and procedure. This includes the pilot test preceding the trials, the video material, the eye tracking software, and the procedure. This subsection will also highlight the reason for using an online eye tracker instead of the eye tracker at the Leiden University Centre of Linguistics (LUCL).

3.4.1 the pilot test

Before the trials with the participant started, the experiment was piloted using a test participant. The pilot was done to check the RealEye software, investigate the format of the data that the software would give, and, most importantly, check whether the post-test questions were clear. Based on the results from the pilot test, there were three main problems that needed to be addressed. First, the results from the pilot test showed that the questions in the post-test could not be made too easy as the participants would be able to answer them correctly regardless of the subtitle availability. Second, it appeared that regardless of the relatively short length of the clip, the test participant still found it difficult to recall the actors' names and specific situations referred to on the post-test. Third, the data from the software indicated that credits in the scenes formed a distraction from the subtitles. This matter will be explained more in-depth under subsection 3.2.3.4. These issues were taken into consideration when designing the actual test. Harder questions were created for the post-test and stills were provided as well to each question to remind participants of the scene and character in question. Moreover, a new section of the episode was chosen that did not contain credits to ensure that participants would not be distracted by the credits.

3.4.2 The videos

This subsection discusses the videos that were used for the experiment. It will discuss the reasons for choosing *Chicago Med*, the episode, fragments, and influence of credits on the eye tracking data.

3.4.2.1 The reasons for choosing Chicago Med

Two different videos from the medical television drama *Chicago Med* were used for this research. Both videos were sourced from the same episode: seasons 3, episode 18, *This is now* (Talbert, Sinclair & Carroll, 2018). This drama was chosen for two main reasons.

- *Chicago Med* is an English drama TV show.
- The language that is used in the show is outside of the colloquial register

This show was chosen because it is in English and the language that is used in the show is outside of the colloquial register. This ensured that the participants would be able to view the videos without the subtitles while the register forced the participants to rely on the subtitles to fully understand the videos.

3.4.2.2 The trailer and the video

Two videos were made: a shorter video that served as a trailer to a longer video that was imported to the eye-tracking software. The implementation of the trailer was of importance, as it would familiarize the participants with the context of the video. This would prevent the participants from searching for visual cues during the first minutes of the eye-tracking session and ensure more consistent and subtitle focused eye tracking results. The first video was recorded from 01:20 to 02:20. During the viewing of this short clip, the eye movements were not tracked.

The second video was recorded from 10:10 to 16:04. This video was used for the eye-tracking experiment. I aimed for a six-minute length, as it would allow for enough material

while also being short enough to analyse. This part and this episode were chosen because many different problems and treatments were discussed. It was important that the clip was not an ongoing sequence about one specific topic; it would then be more challenging to create stochastically independent comprehension questions.

3.4.2.3 The presence of credits

An important factor to take into consideration was the placement of the subtitles. In the pilot test, it was noted that the subtitles were sometimes placed in the upper part of the screen instead of the lower part of the screen. This was done to reserve the lower part of the screen for credits to the director and script writers. Once those credits were gone, the subtitles would return to the lower part of the screen for the next line. The back and forth between the display of subtitles at the upper and lower part of the screen complicated the eye-tracking process because it created two areas on interest, which led to less straightforward results than just having one area of interest. Moreover, the credits are extraneous information that the participants might pick up on but should not focus on. To solve this, a different fragment was chosen that did not contain credits. It must be noted that the short clip that served as a trailer did contain credits. However, no eye movements were tracked during that clip. Therefore, that fragment could still be used.

3.4.3 RealEye

In the white paper published by the company, RealEye is described as “an online platform designed to conduct screen-based webcam eye-tracking research” (Lewandowska, 2020, p. 3). This format allowed users to conduct eye tracking experiments on any given laptop. The advantage of this is that it gave me a wider participant pool. It was now possible to source participants outside of Leiden, as the experiment could be conducted on a laptop instead of the eye tracker at the LUCL lab. The downside to this was the inconsistent and lower eye tracking

quality. Aside from the fact that laptop webcams would give less accurate eye tracking data due to the lower resolution, the use of different laptops with (possibly) different resolutions would give inconsistent eye tracking quality. Nevertheless, the suboptimal eye tracking data was anticipated before the start of the research. The limitations of this data collection method are further deliberated under the Discussion chapter.

3.4.3.1 RealEye vs the LUCL lab eye tracker

RealEye was chosen because it was a more practical eye tracker than the LUCL lab eye tracker. During the trials with the eye tracker at the LUCL lab, several problems that would hinder the research were observed. These problems concerned the experiment building process, trial tests, and administrative problems that lead to data extraction issues. The experiment set up could not properly be timed and the videos would either lag or play at a higher display speed during the test trials. Moreover, due to administrative issues, it was only possible to run the experiment in "dummy mode". As a result of which, it was not possible to extract the eye tracking data.

As a result of the abovementioned problems, it was decided that it would be better to use RealEye. RealEye has been used in other academic research as well (e.g., Federico & Brandimonte, 2019; Federico et al., 2021, Albrektson & Xia, 2021), which supports its reliability for academic research and suitability for this research. The software was accessible and portable, which allowed me to source participant outside of Leiden. Moreover, because the experiments could be conducted on any given laptop, it was also not necessary to reserve time slots at the LUCL lab, which made it possible to conduct experiments at any given moment. These characteristics allowed for an efficient setup of the experiments on the RealEye platform and offered more flexibility to conduct the experiments with the participants. Furthermore, the operating system is very user friendly, because RealEye is a commercial eye tracker and needs to be useable for anyone. The user interface and experiment builder were easy to navigate and

the data extraction of the RealEye eye tracker was very straightforward and without difficulties. Although some time was lost on testing the LUCL lab eye tracker, the simplicity of RealEye allowed me to quickly and efficiently set up an experiment.

3.4.4 The procedure

During the actual trials, the participants were tested in open public spaces of Leiden University, Erasmus University Rotterdam, Tilburg University, and Eindhoven University of Technology. The spaces were chosen based on the lighting to ensure that the eye tracking data would be optimal. Before the experiment, the participants were given a consent form and an information form (Appendix C). These contained general information on the experiment and explained that the participant was free to stop with the experiment at any given moment.

After filling in the consent form, the participant was given a pre-test questionnaire. They were then informed that they would watch a trailer and a short video clip of the medical television drama *Chicago Med*. The trailer and clip were taken from seasons 3, episode 18, *This is now* (01:20-02:20 & 10:10-16:40) (Talbert, Sinclair & Carroll, 2018). The participants were also informed that the clip contained potentially shocking images concerning blood, tragedy, and death. The participants were given a chance to opt out of the experiment. If they did not do so, the participants would be shown the trailer to familiarise themselves with the episode's premise and characters.

After watching the trailer on my laptop, the participants were given a link to RealEye. Using this link, they could conduct the experiment on their own laptop, or on my own laptop in case they did not have a laptop with them. The software first calibrated the participant's eye movements before the clip was shown. Once the participant had finished watching the clip, the participant was given the post-test. When the participant had handed in the questionnaire, the test was concluded, and they were compensated for their time with dainties in the form of chocolate Easter eggs.

3.5 Post-test

This subsection goes over the design of the post-test and binomial distribution, which was used to analyse the statistical significance.

3.5.1 The post-test design

The post-test consisted of 6 different open answer questions. It was also important that we could observe a clear difference in the comprehension scores of the two groups. If the participants can get most answers right without subtitles, it would be difficult to observe an improvement in comprehension, as the scores already are relatively high. Thus, it was imperative to make the questions sufficiently difficult.

The multiple-choice question format proved not to be suitable for the test. The research by Cheng (2004) indicated that participants scored better on multiple-choice close tests and the lowest on open-ended questions. In Cheng's post-test, it was revealed that “guessing played a fairly important part in [answering the questions]. The correct answer they chose did not necessarily mean full comprehension of the spoken stimuli” (p. 549). An open question format was chosen to make the questions more challenging and ensure that the participants relied on comprehension rather than elimination-based guessing. We based the assessment on a detailed answer sheet to check the answers, which offered highlighted keywords and core answer requirements. The questionnaire and answer sheet can be found in Appendix B.

The design of the questions was based on Hao et al.'s (2021) post-test design. Hao et al. created a total of 8 multiple-choice questions: one general comprehension question, three details questions; one inference question; and three questions measuring the participant's understanding of polysemous vocabulary within the context. The post-test of this study only focused on detail questions (e.g., Where will the CT-scans and X-rays be taken?) and general comprehension questions (e.g., What is the solution Dr Choi offers for the sudden influx of patients at the ER?). Due to the straightforward dialogue of the actors, there was no option for

an inference question. Moreover, since we were not investigating the reading of the subtitles on a word-level, we believed it would not be necessary to test the participant's understanding of polysemous vocabulary.

The post-test opened with a general question to enquire whether the participants understood the context of the video. This initial question was followed by six questions that participants could score either 1 or 0 points. Each question was also supplemented with a still of the corresponding scene. The decision was made to do this, as the participants were not completely familiar with the actors, and multiple patients were treated simultaneously. By providing the still, we hoped to have been able to refresh the participants' memory and offer clarity on which patient the question concerned. The test finished with whether the participants had subtitles or not.

3.5.2 The post-test score

Answering the comprehension question only results in two possible outcomes; the answer is either true (1) or false (0). According to Bain and Engelhardt (1992), "a performance of an experiment with only two types of outcomes is called a Bernoulli trial" (p. 91). The post-test consists of six independent Bernoulli trials, which together result in a binomial distribution. A binomial distribution has two parameters: how often you repeat the trial and the chance of succeeding per trial (Bain & Engelhardt, 1992). In this case, the binomial distribution had six trials and the underlying parameter p (the chance of answering a question correctly). This research aimed to investigate whether $p_{subtitles}$ was greater than $p_{nosubtitles}$, i.e. $p_{subtitles} > p_{nosubtitles}$. An important assumption for this binomial model is that the six questions are stochastically independent (Bain & Engelhardt, 1992). In practice, this meant that the outcome of one trial (a participant answering a particular question correctly or incorrectly) would indicate nothing about the outcome of a different trial (the same participant answering a different question

correctly or incorrectly). Moreover, the probability of answering a question correctly must be the same for all six trials.

This section described the materials that were used for the experiment. For this research, a preliminary questionnaire was used to collect basic information about the participants to create a participant profile and substantiate the analysis results. The post-test was created to test the comprehension of the participants. More importantly, it was made to investigate whether the participants with subtitles would answer more questions correctly than the participants without subtitles. Because it was important to observe a difference between the two test groups, the questions needed to be difficult to answer. Moreover, it was important that the participants could not take an elimination-based approach to guess the correct answer. Therefore, open-ended questions were chosen instead of multiple-choice questions. The answers to the post-test were either marked correct or incorrect. This binary distribution meant that the questions followed a Bernoulli trial. For the comprehension score, the six independent trials chained together resulted in a binomial distribution. For the binomial distribution, it is important that each trial is stochastically independent and that the probability of answering a question is the same for all trials.

3.6 Summary

This Method chapter explained the methodology for this research. Section 3.1 detailed the adjustments that had been made to Kruger and Steyn's method. The changes were primarily made to fit the smaller-scale research and the limited time frame. Moreover, Kruger and Steyn investigated subtitle reading. However, because their reading index could not be applied to this research, the fraction of subtitle viewing was investigated instead. Section 3.2 explained the selection criteria for the participants. Section 3.3 described the pre-test questionnaire that was designed to gather data on the viewing habits of the participants. Section 3.4 extensively discussed the materials and procedure. Aside from this, the section also explained that a

commercial online eye tracker was used instead of the eye tracker that was available at the LUCL lab for reasons of practicality. Section 3.5 deliberated on the design of the post-test and the binomial distribution that was used to analyse the statistical significance of the results.

4. Analysis

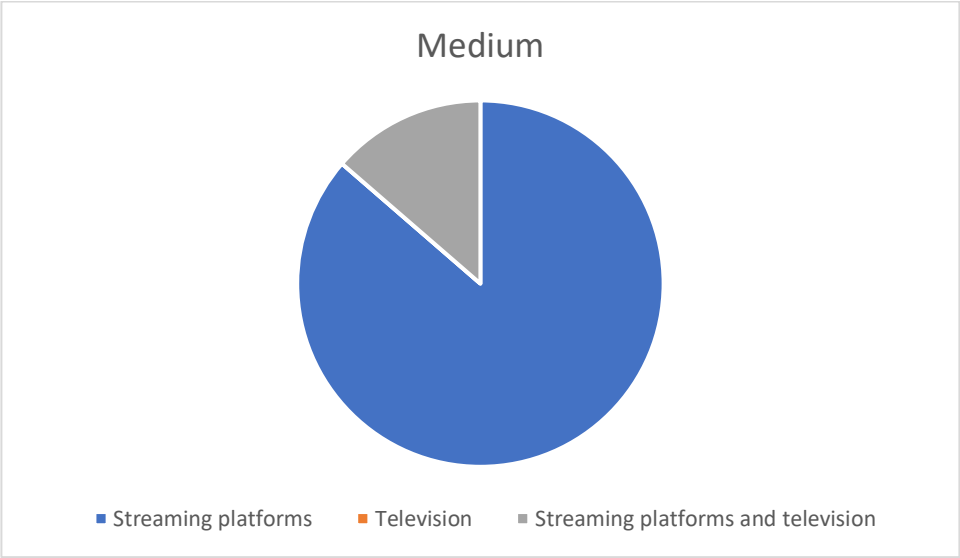
This chapter presents the findings of the pre-test, experiments, and post-test. Section 4.1 presents the results from the pre-test. Section 4.2 focuses on the eye-tracking data. It investigates whether a correlation can be found between looking at the subtitles and answering the questions of the post-test correctly. Finally, section 4.3 shows the results from the post-test and tests the statistical significance of the results. The data was analysed using a Python code that was written by me with help from an outside source (Appendix D).

4.1 The results from the pre-test questionnaire

The pre-test questionnaire was created to gain an understanding of the participants' viewing habits. All participants indicated that they mainly watched shows on streaming platforms. Three participants indicated that they additionally also watched shows on television. These data have been visualised in the pie chart below.

Figure 1

Medium to watch shows

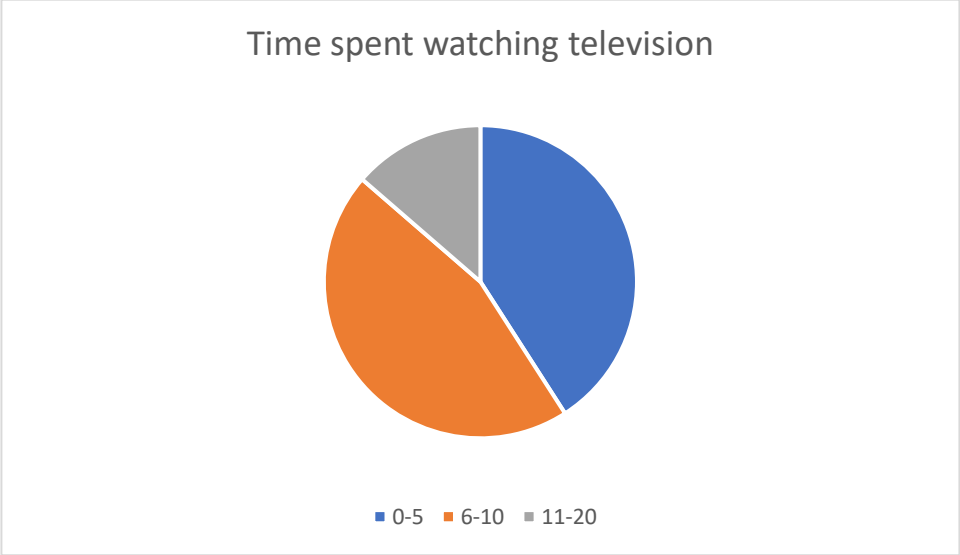


Nineteen participants indicated that they prefer to watch shows on streaming platforms. Although zero participants watched shows on television, three participants indicated that they watched shows both on streaming platforms and television.

Aside from the preferred medium, the participants were also asked how many hours on average they watched shows. That data is collected and presented in the figure below.

Figure 2

The average watch time per week



Nine participants indicated that they watch shows 0-5 hours per week on average. Ten participants indicated that they watch shows 6-10 hours on average per week. Only three participants indicating that they watch 11-20 hours per week on average. However, they did note that this was an extreme and only happened under special circumstances, such as during the holidays.

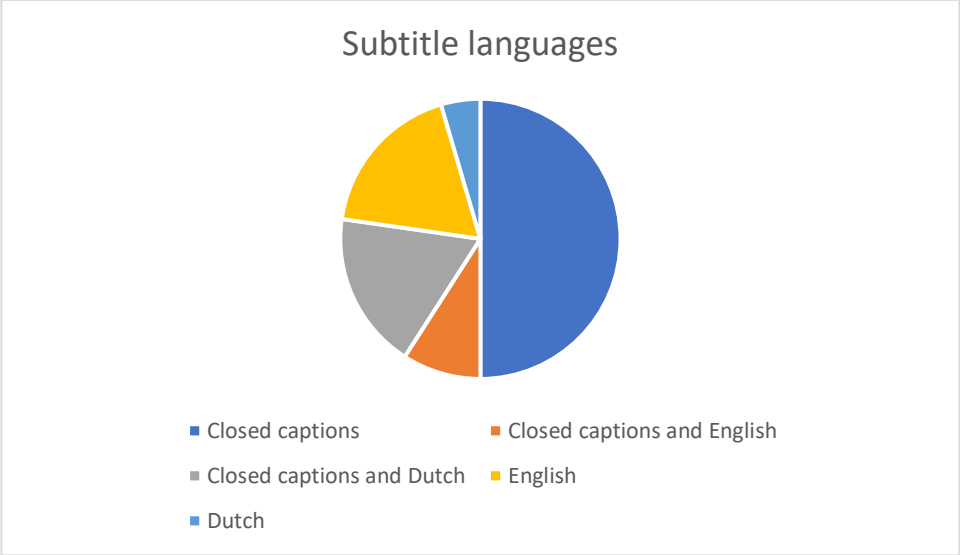
The results from the pre-test questionnaire show that all participants turn on subtitles when watching shows, except for two. The two participants who did not use subtitles indicated that if they did use subtitles, they would turn on closed captions. The audio of the show was oftentimes in either English or the original language of the foreign material (such as Korean for

k-dramas). The participants indicated that for English audio, the subtitles would be set to closed captions or Dutch. For other foreign languages, the subtitles were set to English or Dutch. The subtitle languages have been set out in

Figure ***3*** below.

Figure 3

Subtitle language settings from participants



Eleven participants only used closed captions. Two participants indicated that they used both closed captions and English subtitles. Four participants indicated that they used closed captions and Dutch subtitles. Four participants only used English subtitles and only one participant used Dutch subtitles.

If we take a global look at the results from the pre-test questionnaire, nothing out of the ordinary jumps out. On average, the participants watch shows between 0-10 hours per week. Almost all participants turn on subtitles when they watch a show. The majority turns on closed captions and the participants turn to Dutch subtitles for English shows or English subtitles for other foreign languages (such as Japanese). Nevertheless, data on the participants' viewing habit is important, as they could corroborate findings or reasonings. For example, the high usage

of subtitles could indicate that subtitles are no longer extraneous cognitive load for the participants, since they are already used to them.

4.2 The eye-tracking results from the subtitle group in relation to their performance

Steyn and Kruger (2014) discovered in their research that participants who saw subtitles and also read them performed better on the post-test than the participants who saw the subtitles but did not read them. As was already explained in the Methods, the analysis of this research focuses on subtitle viewing time instead of subtitle reading. Thus, to test whether a higher subtitle viewing time leads to a higher post-test score, two different approaches are taken:

- The percentage of time spent viewing the subtitles is linked to the outcome of each question;
- The performance of the individual participants is investigated.

4.2.1 The percentage of viewing time in relation to the outcome

To investigate whether a longer viewing time results in a higher post-test score, we need the percentage of time that each participant spent viewing the subtitles. We can obtain this fraction by dividing the total time the participants spent viewing the AOI by the total length of the time frame in which the answer to the question was discussed. The time frames can be found in Table 1 below and the fraction of viewing times are presented in

Table ***2***.

Table 1

The time stamps of the fragments that contained the answer to the question

	Q1	Q2	Q3	Q4	Q5	Q6
Time stamp	00:17-00:29	01:45-01:48	02:18-02:23	03:08-03:15	03:28-03:30	04:51-04:53

Table 2

The percentage of time each participant spent viewing the AOI per question

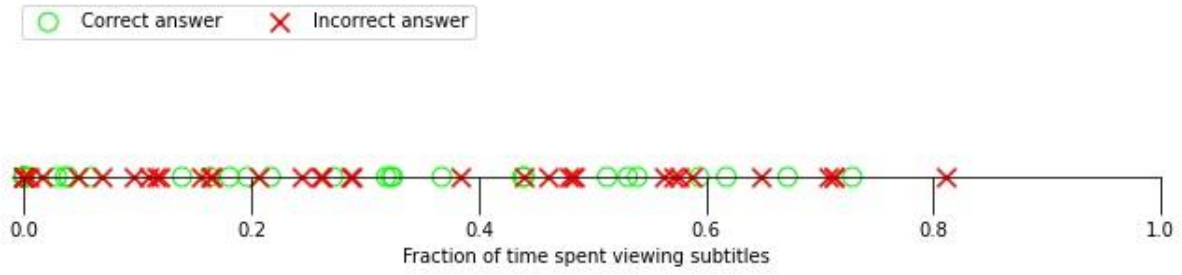
	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Q1	0.18	0.58	0.53	0.05	0.02	0.54	0.12	0.26	0.21	0.14	0.67	0.07
Q2	0.29	0.44	0.49	0.00	0.00	0.71	0.01	0.38	0.56	0.10	0.00	0.17
Q3	0.25	0.16	0.65	0.11	0.01	0.51	0.00	0.48	0.48	0.26	0.71	0.46
Q4	0.20	0.16	0.62	0.04	0.00	0.44	0.00	0.12	0.22	0.06	0.44	0.00
Q5	0.16	0.04	0.73	0.00	0.00	0.81	0.00	0.59	0.27	0.00	0.59	0.71
Q6	0.32	0.00	0.33	0.00	0.00	0.32	0.03	0.29	0.00	0.00	0.37	0.57

These results should be interpreted as percentages of time spent viewing. For example, P1 spent 18% of the time looking at the subtitles for Q1. The segment that discussed the answer for Q1 was 12 seconds long. This means that P1 spent roughly 2 second of those 12 seconds viewing the subtitles.

Figure 4 below presents the distribution of the fractions of time per correct or wrong answer.

Figure 4

Distribution of fractions of time per correct and incorrect answer



Each cross or circle should be interpreted as a time fraction. A green circle signifies a correctly answered question, while a red cross signifies an incorrectly answered question. The location of the observation (either a cross or a circle) indicates the fraction of subtitle viewing time. The line starts at 0.0, which means that, for the observations that fall on this mark, the subtitles were viewed 0% of the time. We can see that, for example, there was a question that was answered

incorrectly while the participant spent about 80% of the time viewing the subtitles. This figure visualizes the classification problem at hand. It shows that the observations of the correct and incorrect answers are too mixed to create a pattern of correct and incorrect answers based on subtitle viewing time. This means that classical classification methods, such as k-nearest neighbours, are of no use.

To demonstrate that the observations are too mixed, we can compare the distributions the observations originate from. A distribution can be defined by its cumulative distribution function (CDF). Thus, the comparison of the distribution of viewing time for correct answers with the distribution of viewing time for incorrect answers is done by comparing the CDF of the distribution for a correct answer (hereafter known as $CDF_{correct}$) to the distribution of an incorrect answer (hereafter known as $CDF_{incorrect}$).

Figure 5

The estimates of the cumulative distribution functions

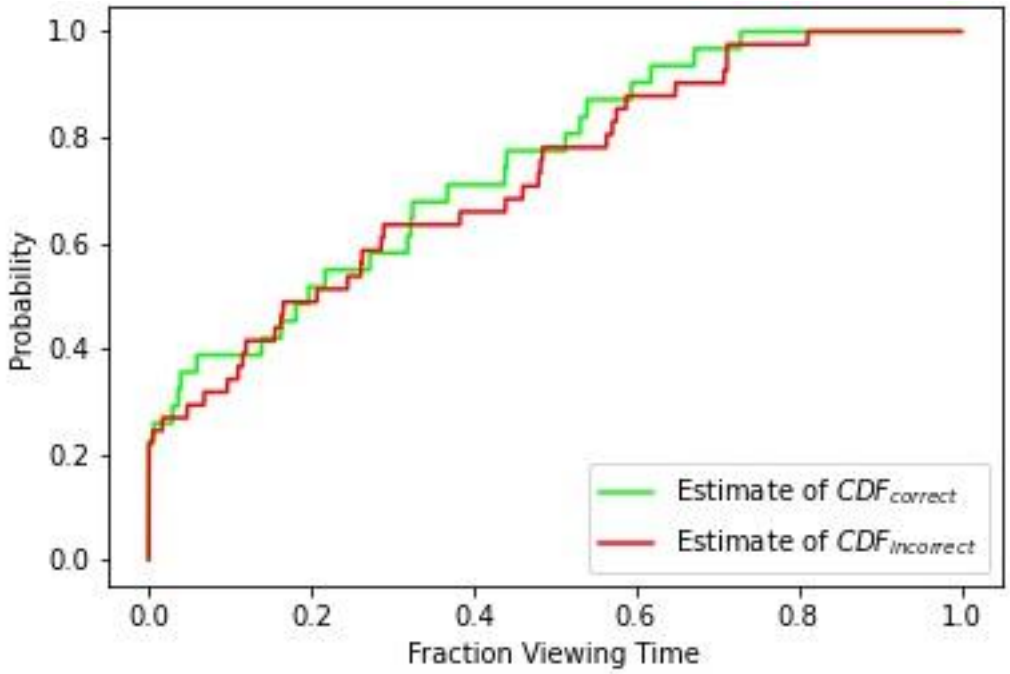


Figure ***5*** shows the estimates of $CDF_{correct}$ with the green line and the red line represents $CDF_{incorrect}$ based on the sample. The x-axis is the fraction of the time, and the y-axis is the probability of the participant looking at the subtitles for the corresponding fraction of viewing time or less. For example, if a participant answered a question incorrectly, there is roughly 40% chance that they looked at the subtitles for at most 15% of the time. We can already see that the two estimates are very close to each other, which already indicates that it will be difficult to differentiate between the estimates based on a given fraction.

To formalize the comparison, the Mann-Whitney U test is performed. The Mann-Whitney U test is a test in non-parametric statistics that shows whether two CDFs are stochastically different (Bain & Engelhardt, 1992). Because this is a non-parametric test, no further assumptions have to be made about the distributions and their sample size. The corresponding hypotheses are:

H_0: $CDF_{correct} = CDF_{incorrect}$

H_1: $CDF_{correct}$ □ $CDF_{incorrect}$.

After performing the Man-Whitney U test, we obtain a U-value of 611.5. For these sample sizes, the U-value corresponds with a p-value of 0.79. this means that the difference measured between the two CDFs, or a larger difference, will occur nearly 80 percent of the time under the null hypothesis. Therefore, there is not enough evidence to reject the null hypothesis in favour of the alternative hypothesis. It is thus not possible to say to which distribution ($CDF_{correct}$ or $CDF_{incorrect}$) an observation belongs solely based on the fraction of viewing time because the distributions are indistinguishable. Translated back to the case of this research, it means it is not possible to confidently classify whether a question was answered correctly or incorrectly based on the subtitle viewing time.

To sum up, nothing can be stated about the difference in distributions of fractions of subtitle viewing time given a correct or incorrect answer. It is thus impossible to create a classifier based on time. I.e., it is not possible to say that an answer is more likely to be correct

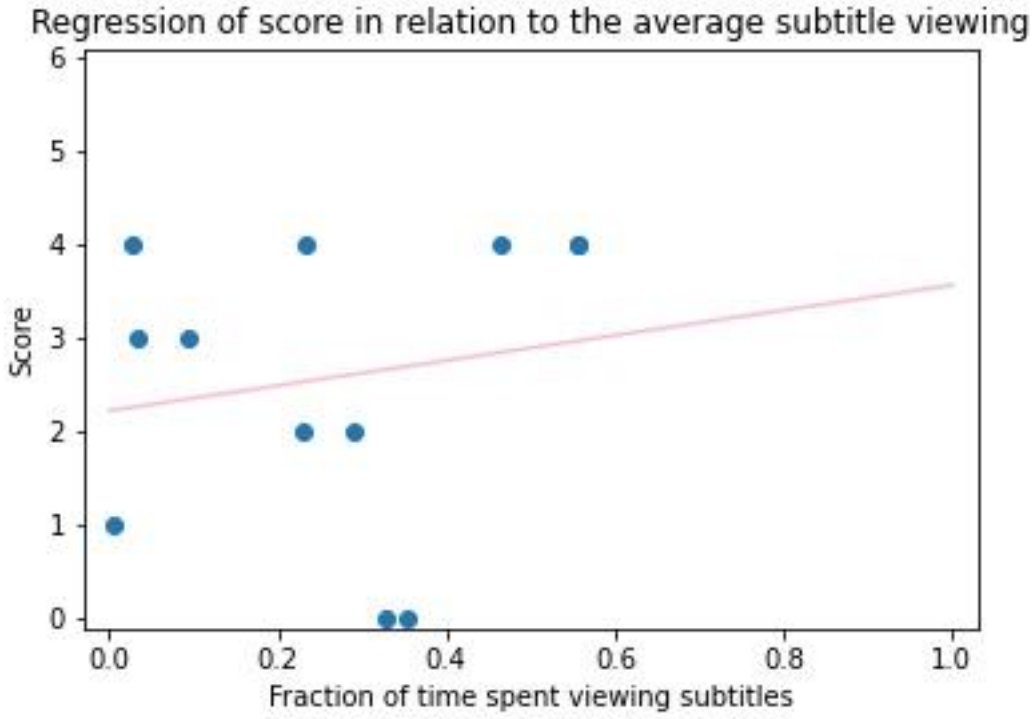
or incorrect based on any given fraction of subtitle viewing time. Nevertheless, this analysis was done based on the assumption that all questions are independent of each other. This is a relatively strong assumption to make, as there are six questions that are answered by the same person. As a robustness check, we relax this assumption. Therefore, the following section takes each individual participant as an observation rather than each question. Note that the observations are still stochastically independent in this case.

4.2.2 The performance of the individual participants

This subsection investigates the performance of each participant who had subtitles to uncover whether subtitle viewing time is correlated to the performance of the participants on the post-test. Different from the test in subsection 4.2.1, the participants are now taken as observations. Figure 6 below shows a regression of the score on the average viewing time.

Figure 6

Regression of the total post-test score in relation to the average subtitle viewing time



Each dot in the regression represents a participant. The y-axis is the number of questions the participant answered correctly. The x-axis is the non-weighted average fraction of time the participant spent viewing the subtitles (the AOI). For example, we can see that there was one participant who had 4 questions correct and spent on average 60% of the time viewing the subtitles. The pink line is the regression line. Simply said, the regression line is a line that runs the closest to the observations.

In Figure 6, we can see that there were some outliers. For example, some participants answered all questions incorrectly despite looking at the subtitles. Nevertheless, the slope of the regression line is 1.341, which is positive. This implies a positive correlation between looking at the subtitles and the score. However, the coefficient is fairly small and has a standard error of 2.46, thus it is not statistically significantly different from 0. Also, the R-squared is only 0.17. I.e., only 17% of the difference in score can be explained by the viewing time of the subtitles.

This subsection investigated the results from the individual participants. Although the slope of the regression line is positive, the R-squared is fairly low. This means that subtitle viewing time is not a good predictor of the total post-test score on its own.

4.2.3 Conclusion of the eye-tracking results

This section showed that it is not possible to prove that a longer viewing time results in a higher post-test score. Subsection 4.2.1 revealed that subtitle viewing time does not indicate anything about whether a question was answered correctly or incorrectly. Subsection 4.2.2 showed that time cannot be used as an indicator for the total score of the post-test. Although the results of the individual participants point towards a slight positive effect of the reading time, the results cannot be held indicative due to the low R-squared and significance level.

4.3 The results from the post-test

The results of the post-test can be found in Table 3, below. Each question has been marked separately with a 1 for a correct answer, and a 0 for an incorrect answer. The participant's total number of correct answers is shown in the penultimate column, and the final column shows whether the participants had subtitles or not. The row at the bottom of the table indicates the total number of participants who correctly answered each question.

Table 3

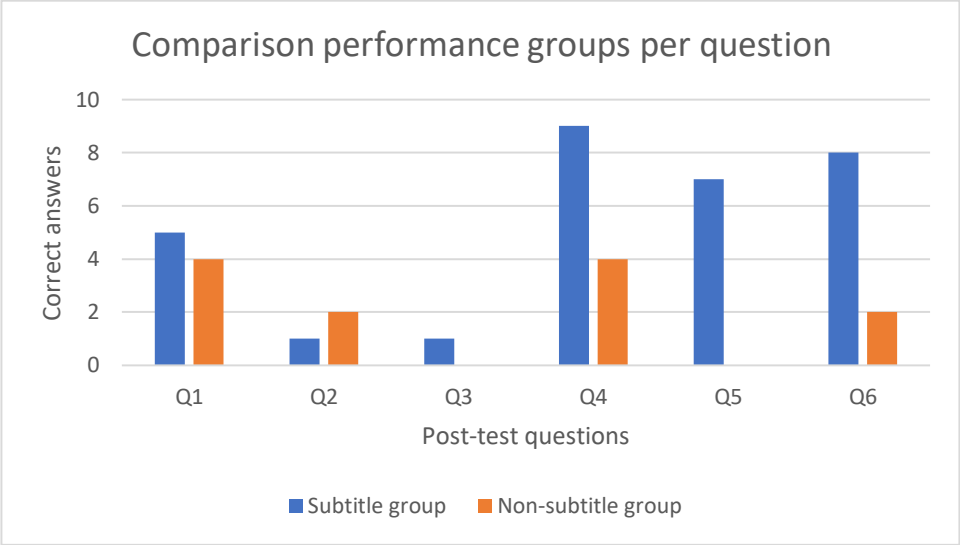
Results from the post-test per participant, per question, total score, total score per question, and subtitle indication

Participants	Q1	Q2	Q3	Q4	Q5	Q6	score:	subtitles
1	1	0	0	1	1	1	4	1
2	0	1	0	1	1	1	4	1
3	1	0	1	1	0	1	4	1
4	1	0	0	1	1	1	4	1
5	0	0	0	0	0	0	0	1
6	0	0	0	0	0	0	0	0
7	0	0	0	1	1	0	2	1
8	0	0	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0
10	0	0	0	1	0	0	1	0
11	1	0	0	1	0	0	2	0
12	1	0	0	1	1	1	4	1
13	0	0	0	1	0	0	1	0
14	1	1	0	1	0	1	4	0
15	0	0	0	0	0	0	0	0
16	0	0	0	1	0	0	1	1
17	0	0	0	1	1	1	3	1
18	1	0	0	1	0	1	3	1
19	1	0	0	0	0	0	1	0
20	0	0	0	0	0	0	0	0
21	0	0	0	0	1	1	2	1
22	1	1	0	0	0	1	3	0
Total:	9	3	1	13	7	10		

The data from the table above has been further processed in the bar chart in Figure *7* below to visualise the difference between the performance of both groups per question.

Figure 7

Comparison of both groups per question



In specific, this bar chart presents the total number of participants per group, per question that answered the question correctly. The blue bar represents the subtitle group whereas the orange group represents the non-subtitle group. By inspecting the bar graph, we may assume the subtitle group outperforms the non-subtitle group and that presence of subtitles results in a higher comprehension score. Subsection 4.3.1 quantifies this by performing a hypothesis test.

4.3.1 Hypothesis test

This subsection performs a hypothesis test to uncover whether the difference between the comprehension score is statistically significant. To do so, we take the following hypotheses2:

$H_0 : p_{subtitles} = p_{nosubtitles}$

$H_1 : p_{subtitles} > p_{nosubtitles}$

where $p_{subtitles}$ denotes the probability of answering a question correctly for the subtitle group. The probability of answering a question correctly for the group without subtitles is denoted by $p_{nosubtitles}$.

Using the data from Table 3, we can calculate the total questions answered correctly by each group and the percentage of questions answered correctly. In total, 72 test questions were asked from the subtitle group. Of those 72 questions, 31 questions were answered correctly. As an unbiased estimator for $p_{subtitles}$ (the probability of answering a question correctly by the subtitle group), we take the percentage of questions answered correctly, which gives us $\hat{p}_{subtitles} \approx 0.43$. The group without subtitles answered 60 questions. Of those 60 questions, 12 were answered correctly. As an unbiased estimator for $p_{nosubtitles}$ (the probability of answering a question correctly by the non-subtitle group), I take the percentage of questions answered correctly, which gives us $\hat{p}_{nosubtitles} = 0.20$.

With this data, test statistic T can be calculated given by:

$$T = \frac{\hat{p}_{subtitles} - \hat{p}_{nosubtitles}}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

where $\hat{p}$ (the probability of the total sample group answering the question correctly) is the unbiased estimator of the parameter of the total sample. Focusing on the part of the questions answered correctly leads us to the following: $\hat{p} = \frac{31+12}{72+60} \approx 0.33$. n_1 is the number of questions asked from the subtitle group and n_2 is the number of questions asked from the group without subtitles.

Under H_0, the distribution of T goes to a standard normal distribution by the Central Limit Theorem as n_1 and n_2 both go to infinity (Bain & Engelhardt, 1992). This means that if H_o is true, the estimators $\hat{p}_{subtitles}$ and $\hat{p}_{nosubtitles}$ would likely be close to each other and the numerator would then be small—the denominator is just a scaling term. Thus, if H_0 is true, we

expect the value of T to be around 0. However, filling in the formula mentioned above, we arrive at T ≈ 2.81. This result is a rather large value of T, which is unlikely to be obtained via a standard normal distribution. The probability of obtaining such or a more extreme result is roughly 0.0024, or less than 0.25%. Given this evidence, we can confidently say that it is unlikely that T truly follows a standard normal distribution. Therefore, we can reject H_0 at a 0.0025 significance level in favour of the alternative hypothesis.

4.3.2 The performance from the groups on answering detail questions

We have ascertained that the presence of subtitles results in a higher comprehension score. However, we can also uncover whether there is a difference between the types of post-test questions in relation to the performance of the groups. Chapter 3 explained that the questions of the post-test score were divided into two categories: detail questions and comprehension questions. Questions 3, 5, and 6 of the post-test were detail questions while question 1, 2, and 4 were comprehension questions. Table 4 below contains an overview of the total score of the groups for the detail questions. We can already note there is disparity between the scores from the detail questions between the two groups.

Table 4

The total of the scores of detail questions per group

	Participants with subtitles	Participants without subtitles
Detail questions	16	2
Total detail questions	36	30

By looking at the scores, we may assume that the participants with subtitles performed better on the detail questions than the participants without subtitles. To test whether the difference between the groups is statistically different, we take the following hypotheses:

$H_0 : pD_{subtitles} = pD_{nosubtitles}$

$H_1 : pD_{subtitles} > pD_{nosubtitles}$

where $pD_{subtitles}$ denotes the probability of answering a detail question correctly with subtitles and $pD_{nosubtitles}$ denotes the probability of answering a detail question correctly.

By repeating the same procedure as described in subsection 4.1.1, we arrive at $T \approx 3.43$. This result is a rather large value of T, which is unlikely to be obtained via a standard normal distribution. The probability of obtaining such or a more extreme result is roughly 0.0003, or less than 0.03%. Given this evidence, we can confidently say that it is unlikely that T truly follows a standard normal distribution. Therefore, we can reject H_0 at a 0.0003 significance level in favour of the alternative hypothesis.

4.3.3 The performance from the groups on answering comprehension questions

The p-value of the test for detail questions showed that the participants with subtitles performed statistically better than the participants without subtitles. Table 5 below shows the results from the comprehension questions. We can see that the difference between groups is much smaller. This subsection investigates whether the difference between the groups is statistically significant.

Table 5

The total of the scores of comprehension questions per group

	Participants with subtitles	Participants without subtitles
Comprehension questions	15	10
Total comprehension questions	36	30

For the test, we take the following hypotheses:

$H_0 : pC_{subtitles} = pC_{nosubtitles}$

$H_1 : pC_{subtitles} > pC_{nosubtitles}$

where $pC_{subtitles}$ denotes the probability of answering a comprehension question correctly with subtitles and $pC_{nosubtitles}$ denotes the probability of answering a comprehension question correctly.

For the comprehension questions, we arrive at $T \approx 0.69$. Contrary to the previous test statistics, the t-value is relatively small. The corresponding p-value is 0.25. This means that the probability of obtaining such or a more extreme result under the null hypothesis is roughly 0.25, or 25%. We do not have enough evidence to confidently reject the null hypothesis.

Thus, if we solely look at the performance scores of both groups, there is reason to believe that the presence of subtitles has a positive effect on the post-test comprehension score. However, there is not enough evidence to conclusively state this. Nevertheless, although we cannot reject the null hypothesis at a meaningful significance level, a larger sample size will be able to clarify this matter.

4.3.4 Conclusion of the post-test results

This section analysed the results from the post-test. Subsection 4.3.1 presented the scores of all participants, and uncovered that the participants who had subtitles performed better on the post-test than the participants without subtitles. Subsection 4.3.2 found that the participants who had subtitles performed better than their counterpart without subtitles on a statistically significant level. Subsection 4.3.3 focused on the performance of the participants on the comprehension questions. Although the participants with subtitles performed better than the participants without subtitles, the result was not statistically significant.

4.4 Conclusion of the analysis

This chapter analysed the results from the pre-test questionnaire, eye tracking data, and post-test. Section 4.1 presented an overview from the pre-test questionnaire. The results showed that most participants spent between 0-10 hours watching shows with subtitles on streaming

platforms. Section 4.2 uncovered that no correlation can be made between the subtitle viewing time and the post-test score. However, section 4.3 shows that participants with subtitles performed better on the post-test than their counterparts without subtitles. This leaves the question as to what could have caused the subtitle group to perform better than the non-subtitle group on the post-test, if not viewing time. The following chapter discusses the answers to the research question and the subquestions. Moreover, it links the findings of this experiment to the literature and offers possible explanations for opposing results.

5. Discussion

This chapter summarizes the aim and findings of this research and relates the findings to the relevant literature. Moreover, this chapter explains the limitations and offers recommendations for future research. The results point towards a positive effect of subtitles on the comprehension of the viewer and hopefully these findings will lead to more research into the study of subtitling. Section 5.1 discusses the results of the subquestions, and section 5.2 answers the research question. Section 5.3 addresses the limitations of this research and recommendations for further research are given in section 5.4.

5.1 The subquestions

The aim of this thesis is to investigate whether subtitles aid comprehension. To answer this question, the following subquestions were set up:

Q1 Does the presence of subtitles result in a higher post-test score?

Q2. Does a higher rate of subtitle viewing lead to a higher post-test score?

Subsection 5.1.1 discusses the findings from subquestion 1 and relates it to the relevant literature. The same will be done for subquestion 2 in subsection 5.1.2.

5.1.1 Does the presence of subtitles result in a higher post-test score?

The results from the post-test indicate that subtitles have a positive effect on the post-test score. In general, the participants with subtitles answered more questions correctly than their counterpart without subtitles, which was further proven to be on a statistically significant level. The presence of subtitles helped the participants to score higher on detail questions than their counterpart without subtitles on a statistically significant level. Subsection 4.1.3 investigated the performance of both groups on comprehension questions. Although the numbers alluded to a positive effect on the comprehension questions, the results were not of statistical significance.

The positive effect of subtitles is contradictory to the findings of Kruger and Steyn's (2014) research. In their research, Kruger and Steyn were not able to find a significant difference between the post-test scores of the participants with subtitles and the participants without subtitles. Nevertheless, the positive effect of subtitles on the performance of the participants is what we could expect based on the Dual Coding Theory. As was explained in the literature review, the DCT hypothesises that language becomes more comprehensible and memorable if the working brain combines the verbal system and visual system. This research provided the subtitles (verbal cues) as a supplement to the audiovisual material. The subtitle group performed better than their counterpart without subtitles, which indicates that subtitles positively influence the viewer's ability to register, recall, and understand information and details.

5.1.2 Does a higher rate of subtitle viewing lead to a higher post-test score?

Contrary to expectations, this research was not able to uncover whether viewing time was an influence on the post-test score. Subsection 4.2.1 compared the fractions of time spent viewing the subtitles for each correct and incorrect answer. The results showed that the distributions from the fractions of viewing time for the correct and incorrect answers were so similar that it was not possible to attribute a given fraction to either distribution. This means that it was not possible to predict the outcome of a question given the fraction of time spent viewing the subtitles. Subsection 4.2.2 investigated the performance of the individual participants with subtitles as a robustness check. The results indicated that the slope of the regression was positive. However, due to the low r-squared, the positive slope could not be held indicative. These results reinforce the conclusion from subsection 4.2.1 that subtitle viewing time is not a good predictor of the total post-test score on its own. Based on the results from this research, subtitle viewing time would not lead to a higher post-test score, as no evidence could be found to suggest otherwise.

Although Kruger and Steyn investigated the influence of subtitle reading rather than the subtitle viewing time, their positive results were taken into consideration. According to Kruger and Steyn, "those who saw the videos with the subtitles and also read the subtitles performed better than those who saw the videos with the subtitles but did not read the subtitles as fully" (p.118). Based on this, it was assumed that the participants who read the subtitles fully also spent a longer time viewing the subtitles in order to fully read the subtitles. Thus, similar results were expected for the outcome of this research; it was expected that a longer viewing time would lead to a higher post-test score. Nevertheless, time could not be found to be an indicator of the post-test results. The results of this research were therefore unexpected. However, it is possible that a number of limitations could have influenced the results obtained.

A source of unreliability could lie in the eye-tracking equipment. Kruger and Steyn used the iView X eye tracker for their research, which has a sampling rate of 50Hz (2014). RealEye is a software that has an average sampling size of 30Hz (Lewandowska, 2020). In essence, "[t]he sampling frequency of an eye tracking system refers to how many times per second the position of the eyes is registered by the eye tracker" (Tobiipro, n.d.). This also means that a higher sampling frequency would allow the eye tracker to better follow the participant's eye movement. The lower level of sampling frequency could have led to noise and uncertainty in the data that could result in statistical insignificance or credibility (Tobiipro, n.d.). This could explain the unexpected results from the viewing time data.

Another possible source of unreliability was the viewing pattern of the participants. It is possible that the participants who answered incorrectly were looking at the subtitles for the answers but could not find the answer before the subtitles disappeared. Unlike Kruger and Steyn, this research did not measure the participant's reading. Therefore, it is not possible to decide whether the subtitles were completely read or only partially.

An additional possible cause for the unexpected result is that the participants were distracted by the subtitles, which caused them to miss the answer to the question. This

hypothesis follows the Cognitive Load Theory, which hypotheses that subtitles would be extraneous cognitive load for the working brain. However, the CLT and the DCT are contradictory to one another in this case. Since the results from the post-test corroborate the DCT, it is unlikely that the CLT could be the cause of the unexpected results. Moreover, almost all participants indicated in the pre-test questionnaire that they turn on subtitles when they watch television shows. This could be indicative of habituation to subtitles, which would lead to a lower cognitive load.

However, the CLT should not be conclusively dismissed. Although the DCT prevails, it is possible that the CLT applies for some participants. The overall performance of the subtitle group is better than their counterpart without subtitles. However, there is still discrepancy between the individual performances of the participants with subtitles. In Figure 6, we can see that the observations are dispersed. Some participants perform as we would expect following the DCT; their subtitle viewing time is high and so is their post-test score. However, some participants perform as we would expect following the DCT; they score low on the post-test despite viewing the subtitles. Therefore, it is possible that the DCT applies for some participants. Moreover, different conditions could also lead to new results. The results might have been different if subtitles were present at the top and at the bottom of the screen. Further research is needed to establish the exact role of the DCT and CLT in subtitle viewing and comprehension.

5.2 The research question

Returning to the question posed at the beginning of this study, it is now possible to state that the presence of subtitles aid viewer comprehension. The evidence from this study points towards the idea that viewers who have subtitles understand and remember the content of the video better than viewers without subtitles, and this research has obtained satisfactory results to be able to confirm this notion. It is, however, unknown what caused the subtitle group to

perform better, as no correlation could be found between time and performance. As mentioned in subsection 5.1.2, there were several limitations that could have caused the unsatisfactory results. Moreover, additional changes could be made to future research in order to gain a better understanding of the mechanisms that cause the viewers with subtitles to perform better than their counterpart without subtitles. Those further recommendations will be laid out in section 5.4.

5.3 Limitations

The research was limited in several ways. First, the method of data collection led to suboptimal eye tracking accuracy. As was briefly explained in the Methods, a commercial eye tracker was used for this research. Although the eye tracker at the LUCL lab was not used for this research, various advantages were observed during the trials. First, the calibration of the eye is conducted manually by the experimenter and the calibration could be conducted in between items shown to the participant, which results in a higher accuracy of calibration and detection. Second, the camera for the LUCL lab eye tracker uses infrared cameras for the detection of the eye. The use of infrared cameras increases the accuracy "[f]or small target images or images near the edges of the screen" (Burton et al., 2014, p.1441). The latter is especially important for research into subtitles, as subtitles are often put either at the top or the bottom of the screen. Finally, as was mentioned in subsection 5.1.2, the average sampling rate of RealEye was relatively low compared to the eye tracker used by Kruger and Steyn. This means that the scan path of the participants was not fully detected by the RealEye eye tracker which resulted in uncertainty in the data. Despite the fact that the LUCL eye tracker offered a much higher level of accuracy, the intricate operating system and experiment set up combined with the limited amount of time made it not possible to bring a proper research with the LUCL lab eye-tracker into fruition. Therefore, the decision was made to opt for a commercial eye tracker and anticipate suboptimal data.

Second, the sample size was too limited for statistical measurements. Although it was possible to prove the statistical significance of subtitles, it was due to the fact that each individual question on the post-test was taken as an observation instead of the entire post-test. This allowed for 132 observations instead of 22. However, for the comprehension questions, there were only 66 observations, which was not enough data to draw a statistically significant conclusion. Moreover, if we look at the regression in figure 3, there are very little data points, as there were only 12 participants with subtitles. The results point towards the idea that the fraction of subtitle viewing time is correlated to the performance of the participants, however, the findings cannot be considered to be representative. The picture is thus still incomplete. Therefore, although the results are promising, they should be validated by a larger sample size.

The final limitation is the question of correlation and causation. In this research, only the correlation between the presence of subtitles and the comprehension of viewers is investigated. According to Antonakis et al. (2010), three conditions must exist in order to prove causality:

- x must precede y temporally
- x must be reliably correlated with y (beyond chance)
- the relation between x and y must not explained y other causes [sic] (p. 1087)

This research does not consider these three conditions and only draws conclusions based on the data. In spite that it is tempting to state that the presence of subtitles causes a higher viewer comprehension, this is not possible. What we can state, however, is that the presence of subtitles is correlated to a higher viewer comprehension. Future research could incorporate these three conditions to gain better insight into the question of causation.

5.4 Recommendations for further research

Further experimental investigations are needed to establish the underlying mechanism that causes the presence of subtitles to be effective for viewer comprehension. One area that has not been explored in this research is the role of subtitles as a trigger to the eye. As outlined

in the literature review, subtitles work as a trigger for the eye. Research by Jensema et al. (2000) has uncovered that the viewers' scan path changes when subtitles appear on screen; the eyes move from the imagery to the subtitles and only return to the middle of the screen after they had finished reading the subtitles. This idea was further corroborated when it was found that viewers still looked at the subtitle despite the audio being in their native language (Szarkowska et al., 2016). It is thus possible that this trigger caused participants to read the subtitles and that, despite the short viewing time, the participants subconsciously absorbed the necessary information.

Another possible area for investigation is the role of subtitles as a focus trigger. After the experiment, one participant commented that they do not pay as much attention to whatever is happening on screen when they do not have subtitles. Kruger et al. (2015) notes that the eyes are drawn to subtitles, as they are identified as a source of information. Therefore, it is possible to hypothesise that the lack of source of information causes the viewers to pay less attention to the events on screen. Attention cannot be measured with an eye tracker. Further research into this area is this necessary to substantiate this hypothesis.

Despite the fact that there are limitations to this research, I believe that this work has shown that watching shows with subtitles is a viable medium for language learning. Although watching television is regarded as a waste of time by some, I would like to argue otherwise.

Works Cited

Albrektson, M. & Xia, C. (2021). *Designing the Navigation to E-services - Improving the Accessibility and Usability to Report a Crime or Loss at the Swedish Police Authority's Website* [Master's thesis, Chalmers University of Technology, University of Gothenburg]. Chalmers Open Digital Repository. https://hdl.handle.net/20.500.12380/303782

Antonakis, J., Bendahan, S., Jacquart, P. & Lalive, R. (2010). On making causal claims: A review and recommendations. *The Leadership Quarterly, 21*, 1068-1120.

Bain, L.J. & Engelhardt, M. (1992). *Introduction to Probability and Mathematical Statistics* (#2). Duxbury.

Burton, L., Albert, W., Flynn., A. (2014). A Comparison of the Performance of Webcam vs. Infrared Eye Tracking Technology. *Proceedings of the Human Factors and Ergonomics Society, 58*(1), pp. 1437-1441.

Castro-Alonso, J.C. & Sweller, J. (2020). The Modality Effect of Cognitive Load Theory. In W. Karwowski, T. Ahram, S. Nazir (Eds.), *Advances in Human Factors in Training, Education, and Learning Sciences* (pp. 75-84). Springer International Publishing.

Cheng, H.S. (2004). A Comparison of Multiple-Choice and Open-Ended Response Formats for the Assessment of Listening Proficiency in English. *Foreign Language Annals, 37*(4).

Chiaro, D. (2012). Audiovisual translation. In C.A. Chapelle (Ed.). *The encyclopedia of applied linguistics* (pp. 1-5). Blackwell Publishing

College voor Toetsen en Examens. (2019). *Moderne Vreemde Talen VWO. Syllabus Centraal Examen 2021 Arabisch, Duits, Engels, Frans, Spaans, Turks.* Retrieved April 1, 2022, from https://www.examenblad.nl/examenstof/syllabus-2021-moderne vreemde/2021/vwo/f=/mvt_vwo_2_versie_2021_def.pdf

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning teaching, assessment*. Cambridge University Press.

Díaz-Cintas, J. (2012). Subtitling. Theory, practice and research. In C. Millán, F. Bartrina (Eds.). *The Routledge Handbook of Translation Studies* (pp. 273-287). Routledge

Díaz-Cintas, J. & Remael, A. *Audiovisual Translation: Subtitling*. Routledge.

Duchowski, A.T. (2017). *Eye Tracking Methodology* (#3). Springer.

Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: the psychology of looking and seeing*. Oxford University Press.

Frederico, G., Osiurak, F. & Brandimonte, M.A. (2021). Hazardous tools: the emergence of reasoning in human tool use. *Psychlogical Research, 85*, 3108-3118.

Frederico, G. & Brandimonte, M.A. (2019). Tool and object affordances: An ecological eye-tracking study. *Brain and Cognition*, *135*, 1-12.

Gifford, C. (Executive Producer). (2000-2019). *Dora the Explorer* [Television show]. Nickelodeon Animated Studio.

Hannula, D. E., Althoff, R. R., Warren, D. E., Riggs, L., Cohen, N. J., & Ryan, J. D. (2010). Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in human neuroscience*, *4*, 1-16.

Harrington, M. (Excutive Producer). (2008-2011). *Ni Hao Kai Lan* [Television show]. Nickelodeon Animation Studio.

Hao, T., Sheng, H., Ardasheva, Y., & Wang, Z. (2021). Effects of Dual Subtitles on Chinese Students' English Listening Comprehension and Vocabulary Learning. *The Asia Pacific Education Researcher*. https://doi.org/10.1007/s40299-021-00601-w

Hervet, G., Guérard, K., Tremblay, S., & Chtourou, M. S. (2011). Is banner blindness genuine? Eye tracking internet text advertising. *Applied cognitive psychology*, *25*(5), 708-716.

Hessels, R. S., & Hooge, I. T. (2019). Eye tracking in developmental cognitive neuroscience The good, the bad and the ugly. *Developmental cognitive neuroscience*, *40*, 1-11.

IMBD. (n.d.). *Chicago Med.* https://www.imdb.com/title/tt4655480/

Ivarsson, J. (2009). The History of Subtitles in Europe. In G.C.F. Fong, K.K.L. Au (Eds). *Dubbing and Subtitling in a World Context.* (pp. 3-12). Chinese University Press.

Jensema, C. J., El Sharkawy, S., Danturthi, R. S., Burch, R., & Hsu, D. (2000). Eye movement patterns of captioned television viewers. *American annals of the deaf*, *145*(3), 275-285.

Karamitroglou, F. (1988). A proposed set of subtitling standards in Europe. *Translation journal*, *2*(2), 1-15.

Kuhl, P.K. (2004). Early Language Acquisition: Cracking the Speech Code. *Neuroscience*, *5*, 831-843.

Kurger, J-L., Krejtz, I., Szarkowska, A. (2015). Subtitles on the Moving Image: An Overview of Eye Tracking Studies. *Refractory: a journal of entertainment media*, *5*, 1-14.

Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, 49(1), 105-120.

Lewandowska, B. (2020). *RealEye Eye-tracking system Technology Whitepaper*. RealEye.

Markham, P.L. (1989). The Effects of Captioned Television Videotapes on the Listening Comprehension of Beginning, Intermediate, and Advanced ESL Students. *Educational Tecnology, 29*(10). 38-41

Mayer, R.E. & Sims, V.K. (1994). For Whom Is a Picture Worth a Thousand Words? Extsions of a Dual-Coding Theory of Multimedia learning. *Journal of Educational Psychology*, *86*(3), 389-401.

Mera, M. (1998). Read my lips: Re-evaluating subtitles and dubbing in Europe. *Links & Letters, 6*, 73-85.

Mayer, R.E. (2002). Multimedia learning. *The Annual Report of Educational Psychology in Japan*, *41*, 27-29

Nation, P. (2007). The four strands. *International Journal of Innovation in Language Learning and Teaching*, *1*, 2–13.

Paivio, A., & Clark, J. M. (2006, September). Dual coding theory and education. In *Draft chapter presented at the conference on Pathways to Literacy Achievement for High Poverty Children at The University of Michigan School of Education*. Citeseer.

Perego, E., Del Missier, F., & Bottiroli, S. (2015). Dubbing versus subtitling in young and older adults: cognitive and evaluative aspects. *Perspectives, Studies in Translatology*, 23(1), 1–21. https://doi.org/10.1080/0907676X.2014.912343

Rijsoverheid, (n.d.). College voor Toetsen en Examens (CvTE). Retrieved April 1, 2022, from https://www.rijksoverheid.nl/contact/contactgids/college-voor-toetsen-enemenscvte

Pujadas, G. & Muñoz, C. (2020). Examining Adolescent EFL Learners' TV Viewing Comprehension Through Captions and Subtitles. *Studies in Second Language Acqusition*, *42*, 551-575

Ranney, L. M., Kowitt, S. D., Queen, T. L., Jarman, K. L., & Goldstein, A. O. (2019). An Eye Tracking Study of Anti-Smoking Messages on Toxic Chemicals in Cigarettes. *International Journal of Environmental Research and Public Health*, *16*(22), 1-12.

Remael, A. (2010). Audiovisual translation. In Y. Gambier, L. van Doorslaer (Eds.), *Handbook of translation studies* (#1) (pp. 12-17). John Benjamins Publishing Company

Richardson, D. C., & Spivey, M. J. (2004). Eye tracking: Characteristics and methods. *Encyclopedia of biomaterials and biomedical engineering*, *3*, 1028-1042.

Rodgers, M.P.H (2013). *English language learning through viewing television: An investigation of comprehension, incidental vocabulary acquisition, lexical coverage, attitudes, and captions* [Doctoral dissertation, Victoria University of Wellington]. Victoria University Wellington. http://researcharchive.vuw.ac.nz/xmlui/bistream/handle/10063/2870/thesis.pdf?sequence=2

Scott, N., Green, C., & Fairley, S. (2016). Investigation of the use of eye tracking to examine tourism advertising effectiveness. *Current Issues in Tourism*, *19*(7), 634-642.

Squid Game subtitles 'change meaning' of Netflix show. (2021, November 4). *BBC*. https://www.bbc.com/news/world-asia-58787264

Stoll, J. (2022). *How many paid subscribers does Netflix have?*. Statista. https://www.statista.com/statistics/250934/quarterly-number-of-netflix-streaming subscribers-wowide/

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science, 12*, 257-285.

Sweller, J. (2005). Implications of Cognitive Load Theory for Multimedia Learning. In R.E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 19-30). Cambridge University Press.

Szarkowska, A., Krejtz, I., Pilipczuk, O., Dutka, L. & Kruger, J.L. (2016). The effects of text editing and subtitle presentation rate on the comprehension and reading patterns of interlingual and intralingual subtitles among deaf, hard of hearing and hearing viewers. *Across Languages and Cultures,* 17(2), 183-204.

Talaván Zanón, N. (2006). Using subtitles to enhance foreign language learning. *Porta Linguarum*, *6*, 41-52

Talbert, E., Sinclair, D. (writers), & Carroll, C.S. (Director). (2018, May 1). This is now [Television series episode]. In Wolf, Dick. (Creator). *Chicago Med.* NBC Studios.

Tobiipro. (n.d.). *Eye tracker sampling frequency*. https://www.tobiipro.com/learn-and support/learn/eye-tracking-essentials/eye tracker-sampling-frequency/

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal, 42*(4), 272-281

Yang, M. (2021, July 19). Having a go: US parents say Peppa Pig is giving their kids British accents. *The Guardian.* https://www.theguardian.com/tv-and-radio/2021/jul/19/peppa-pig-american-kids-british-accents

Appendix A: Pre-test questionnaire

Enquête voorafgaand aan het experiment

De vragen in deze enquête zijn er om data te verzamelen over je opleiding en over je tv-consumptie.

A1. Wat is je leeftijd?

☐☐

A2a. Wat was je hoogst genoten opleiding?

_____

A2b. In welke taal was die opleiding?

_____

A3. Welk middelbaar onderwijsniveau heb je gevolgd?

_____

A4. Kijk je wel eens series of films?

Ja ☐ Nee ☐

A4a. Zo ja, hoeveel uur per week?

Als je het lastige vindt om het in te schatten hoeveel uur je per week kijkt kun je voor jezelf nagaan hoeveel afleveringen je per week kijkt. 1 aflevering van Euphoria/Walking Dead/k-drama is bijvoorbeeld 1 uur. 1 anime-aflevering is ongeveer 20 minuten.

a. 0-5h ☐

b. 6-10h ☐

c. 11-20h ☐

d. 20h+ ☐

A5a. Op welk medium kijk je?

☐ Televisie

☐ Streamingplatform

A5b. Als je op streamingplatforms kijkt, zet je dan de ondertiteling aan?

☐ Ja

☐ Nee

A6. Als je ondertiteling aanzet, in welke taal zet je de ondertiteling en in welke taal is de audio?

Bedankt voor het invullen van de enquête. Je krijgt nu instructies van de coördinator over wat er gaat gebeuren en wat je moet doen. Na het experiment zal je nog een enquête krijgen om in te vullen.

Appendix B: Post-test and Answer sheet

Post-test

Vragenlijst

Je krijgt nu enkele vragen over het fragment dat je zojuist hebt gezien. Q0 en Q7 worden niet meegerekend. Een afbeelding van de bijbehorende scene is bij elke vraag toegevoegd zodat het duidelijk is over welk fragment het gaat. Probeer de vragen zo goed en zo compleet mogelijk te beantwoorden. Als je het antwoord niet weet, kun je de vraag overslaan door er een **X** neer te zetten.

Q0. Wat is er gebeurd waardoor er drukte op de spoedeisende hulp is?

_____

Q1. Wat is de oplossing die Dr. Choi bedenkt voor het aanpakken van de drukte?

_____

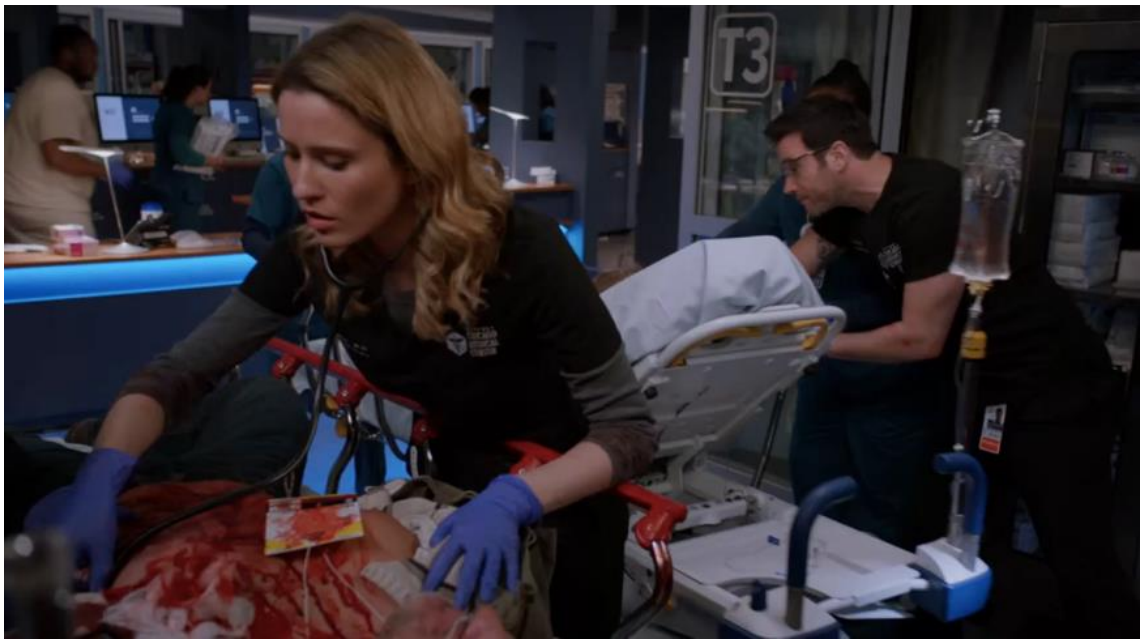
Q2. Waarom moet Dr. Reese het uitstallen van de gevonden voorwerpen overzien?

Q3. Waar worden de CT-scans en Röntgenfoto's genomen?

Q4. Het duurt te lang voordat de zusters bij de medicijnen kunnen. Wat is de oplossing voor het probleem?

Q5. Waardoor is het kind overleden?

Q6. Wat stelt Dr. Bekker vast bij de patiënt?



Q7. Had je ondertiteling?

☐ Ja ☐ Nee

Antwoordmodel / Answer sheet

Vragen worden nagekeken met goed of fout en participanten krijgen 1 of 0 punten per vraag. Participanten krijgen ook punten als hun antwoord neerkomt op het antwoord van het antwoordmodel. In sommige gevallen wordt het ook goed gerekend als de participanten bepaalde woorden noemen i.p.v. een zin. Die woorden zijn onderstreept in het antwoordmodel. Er zijn ook bepaalde woorden die *moeten* worden genoemd. Zonder die woorden, worden er geen punten toegedeeld. Die kernwoorden zijn dikgedrukt aangegeven.

Q0. Wat is er gebeurd waardoor er een drukte op de spoedeisende hulp is? Inkomvraag

A0. Er was een man die rondschoot op een feest / schutter.

I. Participanten moeten aangeven dat er werd geschoten. De locatie maakt niet uit.

Q1. Wat is de oplossing die Dr. Choi bedenkt voor het aanpakken van de drukte?

A1. Ze maken zones/gebieden per behandeling en rouleren de patiënten als een lopende band

I. participanten moeten aangeven dat er zones werden ingericht op basis van behandeling.

Q2. Waarom moet Dr. Reese het uitstallen van de gevonden voorwerpen overzien?

A2. Omdat mensen overweldigd/overwhelmed kunnen zijn.

I. Participanten moeten aangeven dat de slachtoffers overweldigd kunnen worden (door de gevonden voorwerpen).

Q3. Waar worden de CT-scans en Röntgenfoto's genomen?

A3. In het gangpad van/bij de lift. *In the hallway*

I. Het gangpad en de lift moeten genoemd worden

Q4. Wat is de oplossing voor het probleem met de medicatie?

A4. Ze heffen de vingerprintscan op en ze laten de deuren open zodat de zusters makkelijk bij de medicijnen kunnen.
I. Omdat dit gesprek snel werd gehouden mogen de participanten ook antwoorden dat het slot van de apotheek eraf werd gehaald en de deuren opengelaten. Participanten *moeten* noemen dat het slot van de apotheek eraf gaat. De deuren blijven als gevolg protocol open staan.

Q5. Waardoor is het kind overleden?

A5. Vertrappeling.

I. Participanten moeten vertrappeling noemen

Q6. Welke diagnose stelt Dr. Bekker bij de patiënt?

A6. Gat in zijn linkerborst.

I. Participanten moeten antwoorden dat de man een gat heeft in zijn borst. Idealiter wordt er ook aangegeven dat het in de linkerborst is, maar als de participanten de rechterborst aangeven wordt dat niet fout gerekend omdat het gaat om een gat. Als participanten antwoorden dat het een schotwond is, is het ook goed. Als participanten antwoorden dat er een kogel in zijn borst zit, is het fout.

Appendix C: Information and consent form

Information form

Leiden University Centre for Linguistics

Supervisor: Dr T. Reus
Experimentator: Jie Er Lim

Universiteit Leiden

Titel van het onderzoek: The role of subtitles in language learning

Beste deelnemer,

Je bent gevraagd om deel te nemen aan een onderzoek waarmee ik hoop meer te weten te komen over de rol van ondertiteling in taalverwerving.

Inhoud van het onderzoek
Het experiment bestaat uit een sessie van 6 minuten waarbij je oogbewegingen worden gemeten.

Omdat de clips schokkend kunnen zijn voor deelnemers wegens trauma, bloed en overleden patiënten, krijg je de kans om terug te trekken uit het experiment. Mocht je doorgaan en je op een later moment terug willen trekken is dat ook altijd mogelijk. Alle data die je tot dan hebt gegeven zal worden verwijderd.

Voor de start van het onderzoek, krijg je een enquête met enkele vragen over je leeftijd, opleiding, en series kijken. Nadat je de enquête hebt ingevuld, volgt een trailer over de aflevering. Hierbij worden je oogbewegingen *niet* gemeten. Deze trailer is er enkel zodat je een idee hebt waar de clip over zal gaan.

Na de trailer, zal de eye tracker worden opgestart en worden je ogen gekalibreerd. Na de kalibratie zul je een video van 6 minuten kijken. Graag zo stil mogelijk blijven zitten en gelieve niet je ogen samen te knijpen. Nadat het experiment klaar is, zul je nog een vragenlijst krijgen. Beantwoord de vragen zo goed mogelijk, maar voel je niet bezwaard om alle vragen goed te hebben. Als je het antwoord niet weet kun je de vraag overslaan door een **X** in te vullen.

Vergoeding
Voor deelname aan het onderzoek ontvang je als tegemoetkoming van de tijdsbesteding een kleine compensatie in de vorm van een versnapering.

Vrijwilligheid van deelname
Deelname aan dit onderzoek is geheel vrijwillig en vrijblijvend. Dit betekent dat je te allen tijde, zonder opgaaf van reden, kunt besluiten om jouw deelname aan het onderzoek te

beëindigen. Als je besluit het onderzoek te beëindigen heeft dit geen consequenties voor de te ontvangen vergoeding.

Vertrouwelijkheid van informatie

Alle informatie die in het kader van dit onderzoek wordt verzameld, wordt als strikt vertrouwelijk behandeld. Alle gegevens worden in anonieme vorm verwerkt en bewaard. Er zal voor worden gezorgd dat onbevoegden er geen inzage in krijgen en ook dat de gegevens niet tot personen zijn terug te leiden.

Dit onderzoek wordt gecoördineerd door Joey Lim. Indien je vragen hebt over dit onderzoek kun je dat met haar bespreken

Klachten

Indien je vindt dat je onjuist bent geïnformeerd over dit onderzoek, of klachten hebt over de uitvoering of bejegening tijdens dit onderzoek, verdient het aanbeveling dit te bespreken met de onderzoeker of met de coördinator van het onderzoek. Indien je dat niet wilt, of indien dat geen oplossing geeft, kun je ook een klacht indienen bij bestuur van het instituut LUCL. Onderaan vind je de contact gegevens.

Toestemmingsverklaring
Voor deelname aan het onderzoek hebben wij vanzelfsprekend jouw toestemming nodig. Als je bereid bent om mee te doen, kun je dit op het hier bijgevoegde toestemmingsformulier aangeven.

Contact informatie

coördinator /onderzoeker/assistent: Joey Lim
Telefoon: -
E-mail: -

Leiden University Centre for Linguistics (LUCL):
Address: P. N.. van Eyckhof 3, NL-2311 BV, Leiden, Nederland
Telephone: 071-5271662 (Maarit van Gammeren, Institute Manager)
E-mail: lucl@hum.leidenuniv.nl

Consent form

Toestemmingsverklaring

Leiden University Centre for Linguistics

Universiteit Leiden

Supervisor: Dr T. Reus
Experimentator: Jie Er Lim

Titel van het onderzoek: The role of subtitles in language learning

Toestemmingsverklaring (Informed Consent)
Door dit formulier te ondertekenen geef je te kennen de proefpersoneninformatie te hebben gelezen en hebt begrepen. Verder geef je door dit formulier te ondertekenen aan dat je akkoord gaat met de in het informatieformulier beschreven procedures.

Ik heb het informatieformulier gelezen en begrepen en geef toestemming voor deelname aan het onderzoek.

Datum: ……. Plaats: ……

Naam: …… Handtekening: ……...

Appendix D: The Python code

Data Analysis Eye Tracking Master Thesis

June 27, 2022

1 Data analysis

In this file, we analyse all the eye-tracking data.
First, we import the necessary packages.

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy import stats
import math
```

Then, we import the data

[2]:
```
fileLocation = 'D:\Jupyter_Notebooks\Joey\\'
fileName     = 'rawGazesDenoised.csv'
rawData = pd.read_csv(fileLocation + fileName)
```

The goal is to obtain a table containing information on how much each participant looked at the subtitles when the answer to certain questions was discussed. We first create this empty table, so we can store the results in there later on.

[3]:
```
fractionWatchTable = pd.DataFrame(data=None, index = ['Q1', 'Q2', 'Q3', 'Q4',␣
↪'Q5', 'Q6'])
```

To help us with the analysis, we have created a custom function.
This function returns the fraction of time a participant looks at the area of interest (AOI) for a given time frame.
This is done by looking at all the observations and checking whether:
- the time of the observation is in the given time frame; - the X-coordinate of the gaze is within range; - the Y-coordinate of the gaze is within range.

If the observation passess all three checks, we flag it.
Next, we can sum the total time of observations in the AOI for the time frame and return the fraction.

[4]:
```
def rawFixationCheck(dataFrame, beginTime, endTime, xLoc, yLoc):
    timeCheck = np.array(dataFrame['endTime']>beginTime) & np.
↪array(dataFrame['startTime']<endTime) #check the time
```

```
    xCheck    = np.array(dataFrame['xPer']>xLoc[0]) & np.
↪array(dataFrame['xPer']<xLoc[1]) #check the X-location
    yCheck    = np.array(dataFrame['yPer']>yLoc[0]) & np.
↪array(dataFrame['yPer']<yLoc[1]) #check the Y-location
    fixations = timeCheck & xCheck & yCheck #check wheter all three conditions␣
↪are true

    segmentTime = endTime - beginTime #the total time of the segment
    totalTime = 0 #running total of time looked at the AOI
    for count, watchingAOI in enumerate(fixations): #loop over all observations
        if watchingAOI: #the observation is flagged
            totalTime += min(endTime,dataFrame.iloc[count,3]) -␣
↪max(beginTime,dataFrame.iloc[count,2]) #add corresponding time to the␣
↪running total

    fractionWatched = totalTime/segmentTime #get the fraction
    return(fractionWatched)
```

Next, we can loop over all the participants and analyse their eye-tracking data. Unfortunately, the data is currently in unusable form. Thus, we spend some lines converting this into a nice table.

[5]:
```
for participant in rawData.index: #loop through all the participants
    participantTag = rawData['participant_tags'][participant] #store the␣
↪particpant tag
    participantList = [] #create an empty list to store the data for each␣
↪participant. Later this will be added to fractionWatchTable

    fullString = rawData['test_raw_data'][participant] #get all eye-tracking␣
↪data (in string form, which is difficult to analyse)
    fullString = fullString.replace('[','')
    gazeList   = fullString.split(']') #break up into a list containing␣
↪information per timestampe
    gazeList   = gazeList[:-2] #the last two elements are empty due to␣
↪formatting
    gazeTable  = pd.DataFrame(index = range(len(gazeList)), columns = ['xPer',␣
↪'yPer', 'startTime']) #create the table in which all the eye-tracking data␣
↪for the participant will be stored
    #we only need the X-location and Y-location of where the particpant looks.␣
↪We also need the timestamp of when the look initiated
    #The table is currently empty, we add the data next
    for gaze in range(len(gazeList)):
        gazeTable.loc[gaze] = (gazeList[gaze].split(','))[1:4] #add the useful␣
↪data in gazeTable
```

```
    times = (gazeTable['startTime']).tolist() #currently we only have the␣
↪timestamp of when the look initiated, it is also usefull to know when it␣
↪stops
    times = times[1:] + [354000]
    gazeTable['endTime'] = times #add the timestamp of when a look stops (this␣
↪is when the next look is initiated)
    gazeTable = gazeTable.apply(pd.to_numeric) #convert everything to numbers␣
↪to do calculation
    #Now the data is in a nice format and we can start the analysis

    #The answer for the first question is discussed from 0:17 (17000ms) to 0:29␣
↪(29000ms)
    participantList.append(rawFixationCheck(gazeTable, 17000, 29000, [30,70],␣
↪[70,100])) #check what fraction of the time the participant looked at the␣
↪subtitles and add this to the list
    #repeat this for the remaining questions

    #The answer for the second question is discussed from 1:45 (105000ms) to 1:
↪48 (108000ms)
    participantList.append(rawFixationCheck(gazeTable, 105000, 108000, [30,70],␣
↪[70,100]))

    #The answer for the third question is discussed from 2:18 (138000ms) to 2:
↪23 (143000ms)
    participantList.append(rawFixationCheck(gazeTable, 138000, 143000, [30,70],␣
↪[70,100]))

    #The answer for the fourth question is discussed from 3:08 (188000ms) to 3:
↪15 (195000ms)
    participantList.append(rawFixationCheck(gazeTable, 188000, 195000, [30,70],␣
↪[70,100]))

    #The answer for the fifth question is discussed from 3:28 (208000ms) to 3:
↪30 (210000ms)
    participantList.append(rawFixationCheck(gazeTable, 208000, 210000, [30,70],␣
↪[70,100]))

    #The answer for the sixth question is discussed from 4:51 (291000ms) to 4:
↪53 (293000ms)
    participantList.append(rawFixationCheck(gazeTable, 291000, 293000, [30,70],␣
↪[70,100]))

    #now we have a list with how much the participant looks at the subtitles␣
↪when a question is discussed
    fractionWatchTable[participantTag] = participantList #add this to our main␣
↪table
```

```
print(fractionWatchTable) #print the table to look at it
```

```
       12:46     13:25     14:20    14:30   14:45         d         r  \
Q1  0.180750  0.575333  0.531000  0.04725  0.0165  0.539417  0.119500
Q2  0.289000  0.439000  0.484667  0.00000  0.0000  0.708000  0.005667
Q3  0.244800  0.162800  0.648200  0.10980  0.0046  0.512800  0.000000
Q4  0.196571  0.155143  0.617857  0.03900  0.0000  0.438143  0.000000
Q5  0.163000  0.036000  0.728500  0.00000  0.0000  0.811000  0.000000
Q6  0.318500  0.000000  0.324500  0.00000  0.0000  0.322500  0.029000

       15:15     15:30         1         2         3
Q1  0.262500  0.206833  0.139000  0.671667  0.068000
Q2  0.383000  0.563333  0.096667  0.000000  0.165333
Q3  0.479600  0.482200  0.261000  0.711800  0.460200
Q4  0.115429  0.216857  0.059000  0.440286  0.000000
Q5  0.587500  0.272000  0.000000  0.594000  0.712000
Q6  0.286000  0.000000  0.000000  0.367500  0.570500
```

1.1 Linking the percentage of subtitle viewing time to the answers of the comprehension test

We want to link these fraction with the corresponding questions.
By doing this, we have an overview of the time spent on the subtitles of the segment and whether the question was answered correctly or not.
The points per question can be found below:

[6]:
```
t1246 = [1,0,0,1,1,1]
t1325 = [0,0,0,0,1,1]
t1420 = [1,0,0,1,1,1]
t1430 = [0,0,0,1,1,1]
t1445 = [0,0,0,1,0,0]
td    = [1,0,1,1,0,1]
tr    = [0,1,0,1,1,1]
t1515 = [0,0,0,0,0,0]
t1530 = [0,0,0,1,1,0]
t1    = [1,0,0,1,0,1]
t2    = [1,0,0,1,1,1]
t3    = [0,0,0,0,0,0]
allAnswers = [t1246, t1325, t1420, t1430, t1445, td, tr, t1515, t1530, t1, t2,␣
 ↪t3]
```

We create a nested list where the outer list contains 72 inner lists that each corresponding to a question.
Every inner list has two elements: the first one is how long the participant looked at the fraction and the second one is whether the question was answered correctly or not.

[7]:
```
outerList = []
for participant, answers in enumerate(allAnswers):
    for Qnumber, score in enumerate(answers):
        outerList.append([fractionWatchTable.iloc[Qnumber, participant], score])
```

[8]:
```
sumCorrect = 0
nCorrect = 0
sumFalse = 0
nFalse = 0
correctData = []
falseData   = []

for Q in outerList:
    if Q[1] == 1: #if correct
        nCorrect += 1 #count the correct answer
        sumCorrect += Q[0] #add to the total time
        correctData.append(Q[0]) #store the time
    else: #false
        nFalse += 1
        sumFalse += Q[0]
        falseData.append(Q[0])
avgCorrect = sumCorrect/nCorrect #divide by total to get the average
avgFalse = sumFalse/nFalse

print('For correct answers the average fraction of looking at the subtitles is␣
 ↪{a1},\nwhereas for incorrect answers it is {a2}'.format(a1 =␣
 ↪round(avgCorrect,4), a2 = round(avgFalse,4)))
```

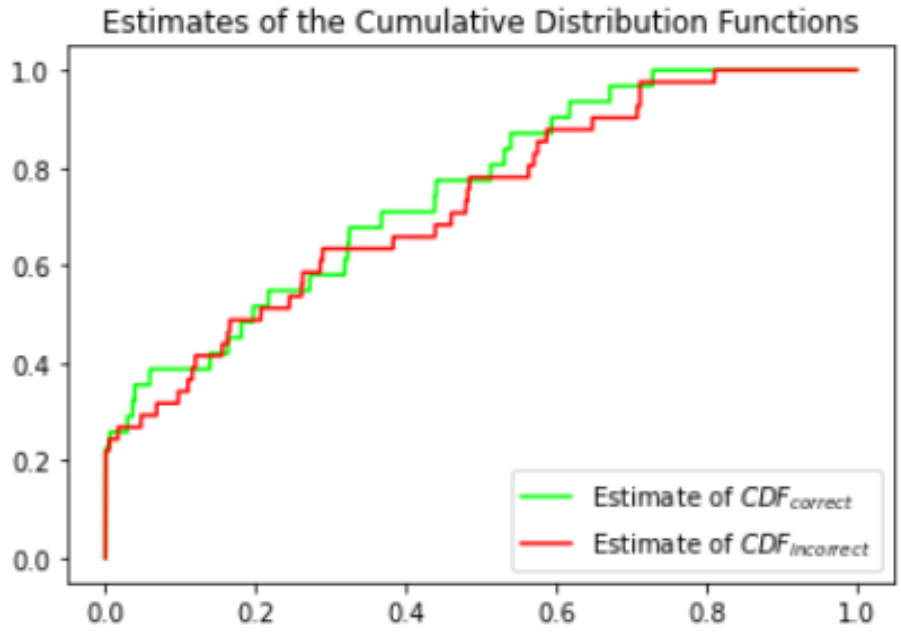
```
For correct answers the average fraction of looking at the subtitles is 0.2498,
whereas for incorrect answers it is 0.2738
```

[9]:
```
stepSize = 100000
xDataCDF = np.linspace(0,1,stepSize)
correctCDF = np.zeros(stepSize)
falseCDF   = np.zeros(stepSize)

for Q in outerList:
    if Q[1] == 1: #if correct
        correctCDF = correctCDF + [int(i) for i in (xDataCDF > Q[0])] #what␣
 ↪values should be raised
    else:
        falseCDF   = falseCDF   + [int(i) for i in (xDataCDF > Q[0])]
correctCDF = correctCDF/nCorrect #normalize
falseCDF   = falseCDF/nFalse     #normalize

plt.plot(xDataCDF, correctCDF, 'lime', label = 'Estimate of $CDF_{correct}$')
plt.plot(xDataCDF, falseCDF, 'red', label = 'Estimate of $CDF_{incorrect}$')
```

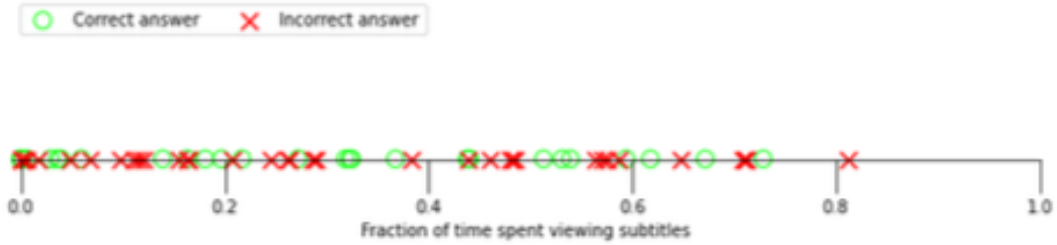
```
plt.legend(loc='lower right')
plt.title('Estimates of the Cumulative Distribution Functions')
plt.savefig('CDFestimates.jpg')
plt.show()
```



[9]:
```
fig, ax = plt.subplots(figsize=(9,3))
plt.scatter(correctData, np.full_like(correctData, 0),␣
 ↪marker='o',s=100,facecolors='none', edgecolor='lime', label='Correct answer')
plt.scatter(falseData, np.full_like(falseData,0), marker='x', s=100,color =␣
 ↪'red', label = 'Incorrect answer')

ax.spines['bottom'].set_position('zero') #make the graph look nice
ax.spines['top'].set_color('none')
ax.spines['right'].set_color('none')
ax.spines['left'].set_color('none')
ax.tick_params(axis='x', length=20)
ax.set_yticks([])  # turn off the yticks

plt.legend(ncol = 2, loc = 'upper left')
plt.xlim([-0.01,1])
plt.tight_layout()
plt.xlabel('Fraction of time spent viewing subtitles')
plt.savefig('MannWhitney.jpg')
```



We can see that there is no clear pattern to whether a question will be answered correctly based on the percentage of time a participant looked at the subtitles.
Furthermore, the average fraction of time spent looking at the subtitles was lower when a question was answered correctly compared to when it was answered incorrectly. Although the average would imply that looking at the subtitles would not help, this is not necessarily the case.
Instead, further analysis is necessary to investigate this relation as Kruger and Steyn's (2014) research showed an adversary result.

[10]:
```
[U,t] = stats.mannwhitneyu(correctData, falseData, alternative='two-sided')
print('The U-value of the Mann-Whitney test is {u}, which has a corresponding␣
 ↪t-value of {t}.'.format(u=U, t = round(t,3)))
```

```
The U-value of the Mann-Whitney test is 611.5, which has a corresponding t-value
of 0.788.
```

This approach did not give us the insight we had hoped for.
Therefore, we will try to analyze the data on the participant level instead of on each individual question level.

1.2 Per participant

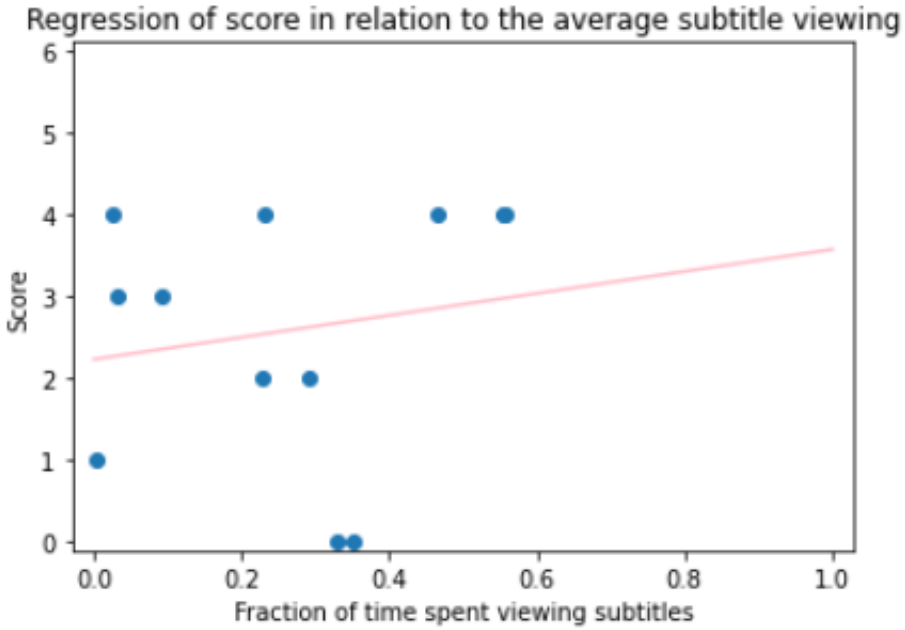
The new plan is to take one participant as an observation instead of a question.
Afterwards, we can fit a regression on the data points.

[11]:
```
partList = []
for i,j in enumerate(allAnswers):
    partList.append([fractionWatchTable.iloc[:,i].mean(), sum(j)])

xData = [i[0] for i in partList]
yData = [j[1] for j in partList]
slope, intercept, r, p, std_err = stats.linregress(xData, yData) #run the␣
 ↪regression
def linReg(x): #function of the line
    return slope*x + intercept
xDataReg = np.linspace(0,1,num=100)
```

```
regression = list(map(linReg, xDataReg))

plt.scatter(xData, yData)
plt.plot(xDataReg, regression, 'pink')
plt.xlim([-0.03,1.03])
plt.ylim([-0.1,6.1])
plt.title('Regression of score in relation to the average subtitle viewing')
plt.xlabel('Fraction of time spent viewing subtitles')
plt.ylabel('Score')
plt.savefig('regression.jpg')
plt.show()
print('The slope of the regression line is {s}, which has a standard error of␣
 ↪{stderr}.'.format(s=round(slope,3), stderr = round(std_err,3)))
print('The R-squared of the regression is {r2}.'.format(r2 = round(r,3)))
```




```
The slope of the regression line is 1.341, which has a standard error of 2.458.
The R-squared of the regression is 0.17.
```

The slope of the regression is positive. This implies that there is a positive correlation between looking at subtitles and the score.
However, the slope is fairly small and the R-squared is only 0.17, thus the correlation is not very strong.
An interpretation of the R-squared is how much of the variation in the score can be explained by the difference in looking at the subtitles.

I.e., only 17% of the difference in score can be explained by viewing time.