Universitat Autònoma de Barcelona

Attention and Dual Coding Theory: an Interaction Model Using Subtitles as a Paradigm

Andreu Oliver Moreno

Tesis doctoral

Departament de Psiclogia Basica i de l’Educació

Attention and Dual Coding Theory: an Interaction Model Using Subtitles as a Paradigm

Andreu Oliver Moreno

Olga Soler Vilageliu

Autor: Andreu Oliver Moreno
Directora: Dra. Olga Soler Vilageliu
Títol Original: Attention and Dual Coding Theory: an Interaction Model Using Subtitles as a Paradigm.
Títol Català: Atenció i Teoria de la Codificació Dual: un model integrador amb subtítols com a paradigma experimental.
Universitat: Universitat Autònoma de Barcelona
Departament: Departament de Psiclogia Basica i de l'Educació
Programa: Doctorat en Psicologia de la Comunicació i Canvi
Any de Diposit: 2017

Signatura de la Directora	Signatura del Doctorand
Dra. Olga Soler Vilageliu	Andreu Oliver Moreno
A Bellaterra dia 27 de Gener de 2017.	A Bellaterra dia 27 de Gener de 2017.

Attention and Dual Coding Theory a Study on the Interaction

Andreu Oliver Moreno

Universitat Autònoma de Barcelona

Departament de Psicologia Bàsica i de l'Educació

To all the family and friends without whom

I could not have reached this point

Abstract

We have conducted two experiments on subtitling and its presentation as part of the HBB4ALL European project. Nowadays, the tendency for content consumption is shifting towards an "on demand" dynamic performed on an array of different devices. With that in mind, not only did we explore the effect the audio language has on the way we process content but also the ways in which it is affected when using different devices to consume content.

In the first experiment we explored the relation between attention, subtitles and information processing. We presented the participants with two sets of clips, one with audio in English and the other with audio in Norwegian in order to explore the effect on the reading of subtitles (in Polish). As predicted by the Dual Coding Theory (Paivio, 1986) multiple channel information processing is found to be a strong facilitator towards better performance in content comprehension. Moreover, attention resource distribution is also found to be altered by the experimental conditions. Resource allocation and demand evaluation mechanisms (Kahneman, 1973) are considered to be the main causes of those differences.

In the second experiment the focus is more applied. A comparison was made between information processing on different devices in order to explore how reading and processing behaviour changes from platform to platform. To assess whether or not there are differences between the different platforms, we presented a set of clips in an unknown language and compared the results of comprehension and visual questionnaires as well as eye tracking readings. No significant differences were found across devices in content comprehension. It is not necessary to adapt subtitles to fit these devices as we are capable of compensating for any differences in their characteristics.

The following general discussion considers the findings of the two experiments to be relevant to the understanding of our attentional and learning system. In both experimental situations, our cognitive processes are able to retrieve the necessary mechanisms to compensate for the differences made in the experimental manipulation.

Keywords: Attention, Dual Coding Theory, Eye Tracking, Subtitles

Resum

Dins l'entorn del projecte europeu HBB4ALL presentem dos experiments relacionats amb el subtitulat. Actualment, estem immersos en un context on el consum dels continguts multimèdia disponibles està migrant cap a una reproducció "a demanda" que es du a terme mitjançant un seguit de dispositius diferents. Amb aquest context en ment, explorem les diferències que la llengua original del contingut genera i les diferències que el consum en diferents dispositius influeixen en la forma que tenim de processar aquest contingut.

Al primer experiment, explorem la relació entre l'atenció, els subtítols i el processament de la informació. Es presenten als participants dos conjunts de clips, un amb l'àudio en anglès (una llengua coneguda pels participants) i l'altre en noruec (una llengua totalment desconeguda) per tal d'explorar l'efecte que es produeix a l'hora de llegir subtítols (en polonès, llengua materna). Tal i com es prediu a la Teoria de la Codificació del Doble Canal (*Dual Coding Theory*) (Paivio, 1986), s'ha trobat que la possibilitat d'accedir a la mateixa informació per per múltiples canals (en el cas que la llengua de l'àudio és coneguda pels participants) facilita en gran mesura la comprensió del contingut. A més, la distribució de l'atenció es veu alterada també en les dues condicions experimentals. La hipòtesi proposada és que el mecanisme de distribució de recursos i el mecanisme d'avaluació de la demanda (Kahneman, 1973), són els principals causants de les diferències trobades en el processament de la informació dels continguts.

El segon experiment és de caire més aplicat. Es duu a terme una comparació entre diferents dispositius per tal d'explorar com la lectura i el processament de la informació es modifica segons la plataforma. Per tal d'avaluar aquestes diferències, es presenten un conjunt de clips en un idioma desconegut pels participants mentre es registra la comprensió mitjançant qüestionaris i el moviment ocular mitjançant eye tracking. No es trobarem diferències significatives entre els dispositius pel que fa a comprensió, pel que es conclou que no sembla ser necessari adaptar els subtítols a les diferents mides de pantalla, posat que el nostre sistema de processament de la informació es capaç de compensar aquesta diferència.

La conseqüent discussió considera que ambdós resultats són rellevants per a la comprensió del nostres sistemes atencionals i d'aprenentatge. En ambdues situacions experimentals el nostre sistema cognitiu es capaç d'activar els mecanismes necessàris per compensar les diferències provocades en les dues condicions experimentals.

Paraules Claus: Atenció, Teoria de la Doble Codificació, Eye Tracking, Subtítols

Index

Glossary

AOI	*Area of Interest*
B&W	*Black and White*
DTC	*Dual Coding Theory*
DTTv	*Digital Terrestrial Television*
HBB4ALL	*Hybrid Broadcast and Broadband for All (European Project)*
HBBTv	*Hybrid Broadcast and Broadband Television*
RT	*Response Time*
VoD	*Video on Demand*

Introduction

To understand the aim of this dissertation we must explain the three main parts of which it is composed. The first part to take into account is the framework through which this dissertation was conceived and financed: the European Project HBB4ALL. This project has come about as a result of the evolution of media platforms, as well as television and its relation with the internet. In the European Union framework, the Hybrid Broadcast Broadband standard is the response to that evolution.

The two cognitive processes of interest in this PhD are related to the individual users on the receiving end as media consumption entities. They are: the attention that consumers pay to the different parts of the media playback and the information they glean from it. To explore in depth these three issues we are going to divide them into three separate sections:

1. The evolution of media platforms, a brief historical overview and the project framework;
2. Introduction of The Dual Coding Theory (DCT), which is the theoretical basis we use to explain how consumers acquire information from media;
3. Introduction of The Attentional Perspective, which, together with the DCT explained in the previous chapter, will help us to understand how our information system captures and processes information.

The two different experiments that form a part of this PhD will follow. The first one relates to the distribution of attention and its relation with the DCT while the second one, refers to the relationship between attention and the evolution of the medium itself.

The evolution of media platforms

Content visualization has changed significantly in the last few years. The technological developments we are seeing define how new generations will consume the content that is produced by the industry. This change is not only related to technical and hardware aspects of the process but also with new platforms, software and technology being made available to the general population. In this day and age, we are no longer limited to what cable TV has to offer or to the visual presentation of this content. Within certain limits, as we will see, we now have complete freedom over where, when and how we want to consume content despite the fact that there are some technical issues that we need to solve.

To understand how we have reached this point, we need to go back to 1926, more than 88 years ago, to the birth of television. The history of this form of media has made TV the king of communication, becoming one of the world's largest businesses. Nevertheless, the major changes television has undergone can be counted on one hand: The transition from B&W to colour, satellite broadcasting, digital TV (that led to flat-screen TVs) and the latest, Digital Terrestrial Television (DTT) the uptake of which is illustrated by the fact that the number of countries which have switched from analogue to digital signal increased from only 4 in 2008 to 37 in 2015 (International Television Expert Group, 2013).

One could say that this lack of change is what has made the TV so broadly accepted, but nowadays, in the communications era, television has been cast aside and is the least connective of all devices. Having said that, it is still the most used media device, with more than 33 million potential viewers and an average use of 3h 12 minutes a day with countries such as the United States having an average consumption of 4h 35 minutes a day (International Television Expert Group, 2013).

In the 70s teletext was the first attempt to expand the television's applications (Polder & Parker, 1985). But it wasn't until the new century that the internet as we know it became integrated with the television. Nowadays a European standard on hybrid television is being developed: HbbTv. This technology allows users to merge content transmitted from the regular cable television with information transmitted separately through the internet. As the consortium describes it on their web site: "The HbbTv specification was developed by industry leaders to effectively manage the rapidly increasing amount of available content targeted at today's end consumer. It is based on elements of existing standards and web technologies including OIPF (Open IPTV Forum), CEA-2014 (CE-HTML), W3C (HTML etc.) and DVB Application Signalling Specification (ETSI TS 102 809) and DASH" (HbbTv.org Overview, n.d).

The implications of this technology are mainly related to the capability of the user to personalize all their content. On one hand, this allows Video on Demand (VoD) to be embedded in the same platform as live or aired television giving the user the freedom to choose what they want to see and when. On the other hand, the how comes into play. This hybrid technology allows users to personalise how they are going to watch certain aspects of the content. They are going to be able to change audio language (option already available in DTT), personalise the subtitles and some accessibility aspects, along with other options.

To explain the evolution, we need to consider that parallel to all those advances, a new generation of devices have emerged as well. When it comes to hardware, PCs, mobile phones and tablets have evolved and become more multi-functional, mainly due to huge advances in internet capability and coverage. The first iPhone was launched in June 2007 (Vincent, 2017) and the first iPad in 2010 (Kastrenakes, 2015). Ever since those launches, their functionality has only increased to the point where they set the benchmark for this type of device.

Consequently, the media production market has been in a process of constant adaptation to the medium. The rapid development and adoption of those technologies have pushed all the sectors into an innovation race (Cook, C. I. 2014). The evolution of television and the emergence of a wide range of satellite devices have presented a challenge as well as an opportunity for the entire industry to access a different playground. As Cook (2014) describes it: “What once was an activity restricted to one screen and limited by time, place, and content is now and continues to be an activity increasing in options and possibilities. However, the television industry has been slow to respond to the changes” (p. 3).

The behavioural changes that all this implies can be seen in the younger generations. There is a tendency, constrained of course by availability, to consume content on demand. People can no longer wait to see the show they want at a certain time, on a certain channel and dubbed. On the contrary, they want to watch it online on any device they have to hand when they have the time or the will to do it (Cook, 2014). This is a result, as we said, of the technological advances, but there is also trending necessity to be able to watch the shows as soon as they air (Cook, 2014).

In order to put what we have described into numbers, we can look at the report from “Asociación para la Investigación de Medios de Comunicación” (AIMC, 2016). which collects user information and behaviour related to all media platforms. This association has been conducting this study periodically since 1997, giving us a broad perspective of how content consumption and demand have evolved in Spain since then.

In Graph 1 we can see such evolution in the percentage of penetration between television and internet. The first impression that we can draw from the data is the quick adoption of the internet, which has now reached 69.9% of the audience. Before making any judgment on this, we must take the technological and knowledge barrier associated with the internet for the older generations into consideration. This means that this evolution will continue in relation to the evolution of the population whereas the television has already jumped this generational barrier. If we look at the penetration segmented by ages, we can clearly see this effect: TV is watched by all generations. In the Graph 2 we can see how the internet is predominant from ages 14 to 34, generations that were raised with easy access to the internet. Finally, in 2016 the television is at an all-time low (within the range of this study) with 88.3% penetration.

This data helps us understand and quantify the arguments raised earlier. Content consumption is shifting its bases towards an entirely different spot. This is the context in which HbbTv has emerged. This hybrid technology is the adaptation the television industry is undergoing to bridge the gap between certain ages. It merges different informational channels (broadband Tv and internet connection) in the same box. As a consequence, we will be able to have more information delivered to us as subtitles and different audio channels; it will also lead to a content shift from cable to the internet, meaning access at any time on any device. These are the two topics (subtitling and diversity of devices) that will be the scope of this thesis, in the framework of the European project HBB4ALL.

Owing to this flexibility in access to the content, more attention is being paid to accessibility of this content. Whether it is because of hearing impairments or because of language barriers, subtitles have become of more relevance than ever (Perego, Laskowska, Matamala, Remael, Robert, Szarkowska, Bottiroli, 2016). Nowadays, due to recent technological breakthroughs, as mentioned previously, the demand, consumption and availability of subtitles have increased considerably.

Such technological advances have not only given users access to certain previously unavailable content but also fuelled the ongoing debate over subtitling vs dubbing (see Perego et al. 2016). In Spain, the tradition is to dub films. All the content broadcast on

TV or shown in cinemas has gone through the process of being dubbed into Spanish. Recently although, we have begun to see original version films in some of the more commercial movie theaters. Moreover, with the transition to the DTT technology, all televisions in Spain now come equipped with the options of switching the language to the original audio recording and activating subtitles.

A switch to subtitling would also have economic implications. According to the MCG report in 2007, big market countries such as Spain, France or Italy, spend on average 34,900€ on dubbing a 90-minute film. Whereas subtitling the same film would cost around 5,300€.

We are now facing the evolution of different media at different paces. This evolution is giving rise to a behavioural change in younger generations. The range of content and information at their disposal, the timeframe in which they are able to access it, the way they process it and the medium they use to watch it are all affected by this shift. We are going to focus on two aspects of this far-reaching change in media consumption: the language possibilities that this opens and the concerns related to the support people use to watch specific content. We will attempt to gain an insight into the cognitive processes that those changes may lead to in an ecological environment.

This situation we have described outlines the growing importance of subtitles in Spain as in many countries where dubbing is typical. However, there is still a cultural barrier to cross (Perego et al, 2016). This thesis is centered around two questions:

- How do subtitles affect the way we distribute our attention between scene and captions and what information do we glean from them?
- How is subtitled content affected by viewing it on the aforementioned array of devices?

In order to answer those questions, we must first attempt to understand some of the theories which have been developed in order to offer an explanation of the cognitive processes that can be applied to reading subtitles.

Double Channel Theory

As mentioned above, in this section we will discuss one of the psychological foundations of this study. The Dual Coding Theory was first introduced by Paivio in 1986 in the book "Mental Representations: A dual coding approach". Paivio's theory is based on the notion that redundant information improves content comprehension. In other words, complementing the information of any set content with additional related input or another source of information allows people to better comprehend the meaning of the content.

According to Paivio (1986) this redundant information proposed in the DCT is processed by two different subsystems extracted from Morton's Logogen Model (1969): A verbal system specialized in dealing directly with language and a nonverbal system for nonlinguistic stimuli. Each one of them have specific representations: Logogens for the language system and Imagens for the nonverbal system. As we can see in the Figure 1 those systems are also connected to the response output. These two systems work in connection with each other. As Paivio (2006) states, “the language system [for instance] is a necessary player in all “language games” but it is sufficient in only a few”. One system then dominates in certain tasks while the other dominates in others.

Going back to Morton’s theory, these sets processing information are defined as a “device which accepts information from the sensory analysis mechanisms concerning the properties of linguistic stimuli and from context-producing mechanisms” (Morton, 1969 p. 166). This concept has evolved from the original theory to adapt to new evidence and to extend to other sensorial information creating modality-specific and response-specific logogens. At the same time, the concept of imagen represents, as Paivio (1986) states, “representations from which mental images are generated under appropriate conditions” (p. 59). For instance, each letter on a keyboard and the components that create each key form part of a mental representation in itself and also, belong, in an organised manner to a larger, structured, representation as part of a computer.

The key factor for the DCT relies on the communication carried out between the two representational systems. Certain aspects of a mental concept can elicit representations of the other, the most common example is naming, an object representation (imagen) can elicit the word (logogen) of the object being represented. It’s important to remark thought that even if these situations are present regularly, both systems work in parallel most of the time.

To understand the origin of the Dual Coding Theory (DCT) Paivio (2006) establishes its roots in the evolutions of human epistemology. The relationship between language, imagery and knowledge has shifted according to the evolution of the species. From a pure imagery representation of the known concepts, we rotated towards a balanced distribution where language gained the same relevance as the visual representations.

The *Orbis Sensualium Oictus* ("The world explained in pictures") was one of the first representations of the combination between language and imagery. From there, as Paivio states, it is not much of a conceptual stretch to start thinking about the start of a dual coding system.

The author of the *Orbis*, Jan Amos Comenius, already had an idea of how and why this combination was necessary: "things are essential, words only accidental; things are the body, words but the garment; things are the kernel, words the shell and husk. Both should be present to the intellect at the same time, but particularly the things, since they are as much objects of understanding as is language" (Comenius, 1896 translation, p. 267; cited in Paivio 2006).

From an ontogenetic perspective, this evolution is replicated. As a child we progressively generate cognitive representations based on our experience with physical objects and specific events. These representations are intrinsically nonverbal and are related to each other. Throughout the language acquisition process, we link the new logogenic representations we learn with the existing imagery to generate a dual coding scheme. This system and the relationship between both paths grows in complexity as we learn. From strict referential lexic we move to abstract representations, syntax and grammar until at some point, language shifts from concrete representation to become an independent medium in itself becoming relatively autonomous of the physical world.

Subtitles and Dual Coding Theory

The developmental theory underlying the DCT has a clear connection to an educational application. If we acknowledge the role of a duality in the codification of the information we process, we can thereby apply it in real life and at the same time, assess the effects of this channels while we try to understand how do they work.

Mayer and Moreno (1998) contributed to the Dual Coding Theory with evidence of the learning improvements and advantages of a dual information delivery system based on different sensorial working memories. In the Figure 2 we can see their proposed model for multimedia learning.

To test this model, they tested students understanding of a short animation accompanied by text or by an auditory narration. In the group that only had visual stimulation (text and images) the visual working memory was likely to become overloaded while in the auditory group (audio and images) the fact that the information was delivered via two different channels must have helped the students to better comprehend the information. In other words, Mayer and Moreno (1998) hypostatize that in the text group, "students may not have been able to generate as many connections between corresponding pictorial and verbal information" (p. 314) as the audio group did.

The results they obtained back up this hypothesis. The group with an audio presentation performed better in the three comprehension tests (retention, matching image and names and transfer, as well as problems solving based on the information learned) and this difference remains stable in two different experiments with different clips. Mayer and Moreno's findings are also supported by another real life situation that has also been linked to Paivio's Dual Coding Theory: subtitles. (for a comprehensive overview of the impact of subtitles on foreign language learning see Vanderplank, 1988, or Talaván, 2010).

According to the Oxford Dictionary (n.d), subtitles are "captions displayed at the bottom of a cinema or television screen that translate or transcribe the dialogue or narrative". In subtitle processing, the cognitive effect has been found to be very efficient as it relies on this redundancy (see Gottlieb, 1998; Tomaszkiewicz, 2006). According to Diao, Chandler and Sweller (2007), "the redundancy effect occurs when the same information is presented [...] in different forms", which requires people "to mentally coordinate the multiple forms" (p. 239). D'Ydewalle and Gielen (1992) argue that even though different information sources are "partly redundant", yet they "do not contradict but rather supplement one another" (p. 415).

If we go back to the Mayer and Moreno's (1998) working memory schema, subtitles represent extra input for the visual channel. In the processing of subtitled content, the visual system has to decode two different sets of information, on the scene (understood as what happens in the visual narrative of the clip content) and the subtitles that connect the visual channel with the auditory channel (or not if the audio language is not known by the participant).

From this increase in complexity comes the controversy of subtitles. On the one hand, the addition of subtitles to a video has been thought to increase extraneous cognitive load (Brünken, Plass & Leutner, 2003; Paas & Van Merriënboer, 1993; Paas, Van Merriënboer & Adam, 1994). According to this view, subtitles are "detrimental to learning" (Kruger, Esté & Matthew, 2014, p. 3). However, subtitles are considered to provide viewers with useful visual support (Paas, Tuovinen, Tabbers & Van Gerven, 2003).

By analysing self-reported tests such as the one carried out by Vanderplank (1988) we can see that the variability in people's awareness of subtitles is huge. In his study exchange students were asked to watch at least one hour of subtitled television and give feedback about the experience. On the one hand, as stated in Vanderplank's (1988) report "subjects report that they were conscious of learning a great deal of language from the programs watched" (p. 275) but on the other, some "felt that they would have listened harder without the text, and would have to try and improve their listening ability if the text had not been present" (Vanderplank, 1988, p. 275) and some even reported "text dependencies or of overload channel capacity caused by the addition of the text" (Vanderplank, 1988, p. 276).

Apart from that controversy, other kinds of self-reported statements are of much interest. For instance, he also states that "all subjects reported that they had developed techniques and strategies for minimizing distraction and maximizing the usefulness of the text" (Vanderplank, 1988, p. 276) or the fact that after the feedback sessions, exercises and class activities, Vanderplank's reports an improvement in the participant's language knowledge.

We need to bear in mind though that all those results are based on self-reported tests. However, those findings lead to experimental studies like the one conducted by Hyati and Mohmedi (2011). This paper, published in the British Journal of Educational

Technology, examines the effect of different kinds of subtitles on comprehension levels. In the context of English as a Foreign Language, researchers examined the effect of intra (English audio and English subtitles) and inter (English audio and Persian subtitles) subtitles and their difference when compared to a situation with no subtitles. Results show that performance in a comprehension questionnaire improves when the subtitles are presented in intra-lingual modality, followed by inter-lingual modality. Both groups subjected to multi-channeled information achieved better results in the comprehension test than those who had no textual help.

In their classic form, subtitles contain a translation of a foreign film's dialogue into another language. This is known as interlingual subtitles (Díaz Cintas & Remael 2007; Ivarsson & Carroll, 1998). Another common type is intralingual subtitling, also known as Closed Captioning (Robson, 2004), where the language of the soundtrack is the same as the language of the subtitles, as is frequently the case with subtitling for the deaf and hard of hearing (Neves, 2008). Standard subtitling uses the viewers' mother tongue while the film dialogue is in a foreign language. In reversed subtitling, the film dialogue is in the viewers' native language and subtitles are displayed in a foreign language (Díaz Cintas & Fernández Cruz 2008; Talaván & Rodríguez-Arancón, 2014).

In Hyati and Mohmedi's (2011) study however, some information regarding the way the comprehension questionnaire was conducted and the language in which it was administered is missing. Moreover, we are again faced with an inter-subject study that, given the nature of this experiment, may influence the results. Information on the visual processing of the scene is also missing. We understand that the main goal of the study was to focus on the learning process, but this is not an ecological situation. Participants might have ignored the fact that they were watching a clip in order to improve their performance in the test and simply focused on reading the subtitles. This text-dependency, in a real situation, would have caused a loss of information about the narrative of the film, which would probably have made it impossible to understand completely, and also pointless to watch.

In contrast, the aforementioned self-reported study by Vanderplank (1988) Diao, Chandler, and Sweller, (2007) focused on understanding whether or not text-dependency can affect the participant's ability to understand the auditory signal. In their study published in the American Journal of Psychology "The effect of written text on comprehension of spoken English as a foreign language", they tested the hypothesis that written text had a detrimental effect on the ability of learners of English as a second language to develop skills to better understand auditory only stimuli. Diao, Chandler and Sweller (2007) raise a valid point regarding cognitive load on multiple channel information processing. According to them a redundancy effect occurs when information is presented in different forms requiring a synchronization effort to match information from different channels to each other.

However, results show that the self-reported cognitive load when processing information delivered through multiple channels is lower compared to mono channel. Participants that were shown scripts or subtitles of the auditory delivered content reported lower cognitive load scores than the participants that did the tests only with the auditory content. Nevertheless, this effect was reversed when they removed the written aids from the two groups that had used it. At that point, people who had been using the scripts and subtitles reported a higher cognitive load when they had to face auditory stimuli without that help, whereas people that did all the exercises without that help reported a lower cognitive load with the same stimuli.

Finally, when reading Diao, Chandler and Sweller (2007) study we have to take some questionable methodological decisions into consideration. For instance, the presentation of the clips was neither aleatory nor controlled meaning that the content of the stimuli, as we are able to see in the second experiment 3rd stimuli, affected the results and scores reported. The design is inter not intra-participant, meaning that personal variables, such as their level of English proficiency which is only controlled by the class they are attending, are likely to affect the different scores. Moreover, no bases for the self-reported cognitive load scale are established leaving the participants' understanding of the concept as the key factor in their correct reporting. Finally, the construction of the comprehension test is not transparent and is based only on the textual information whereas if they had forced the participants to pay attention to the visual content, it would have replicated a more realistic scenario. In the situation as it is presented, students could only have been focusing on text representing a scenario (audio-text and no image), which is rarely present in real life.

Another line of research related to this is the language or vocabulary learning paradigm. Bisson, Van Heuven, Conklin and Tunney (2014) tried to take a holistic approach to the matter with little success. They tried to compare vocabulary acquisition by comparing different kinds of subtitle presentation. Intralingual subtitles (Dutch audio Dutch subtitles), interlingual subtitles (Dutch audio English subtitles), a reverse subtitles condition (English audio Dutch subtitles) and a control condition (Dutch audio without subtitles). In an ANOVA comparing the results of a vocabulary test after the viewing of the clips, they found no main effect generated by the subtitle condition ($F_{(4, 49)} < 1$). So far we have established the relevance subtitles have from an educational and learning based perspective. From a broader perspective, we need to consider the cognitive processes behind subtitle processing. That is why (as we pointed out several times) there is an element missing from the equation we are describing: the image processing.

Apart from the complexity caused by language processing on the auditory channel and reading on the visual channel, we need to take screen processing into consideration. In doing so, we are incorporating an overlapping problem that has been completely omitted by many of the studies outlined so far. Based on Kahneman's (1973) attention theory, we could say that this overlapping between the subtitles and the scene is resolved by assigning different amount of resources to each one of the tasks.

Therefore, on the one hand we now have dual information coding and the comprehension and learning enhancement related to the auditory channel that it entails. And on the other hand, we have the overlapping information that needs to be processed on the visual channel at the same time. Subtitled text is required in order to comprehend the auditory input we receive whereas visual scenic information is necessary to understand and enjoy the content we are viewing. This overlap needs to be resolved by our attention system in the most efficient way possible to ensure there is no loss of information.

The attentional perspective

In this context a question arises: how do subtitles affect the process of understanding and viewing played content? The underlying factor in all the papers we have studied is attention. As a modulator of other cognitive functions it is a very complex mechanism. It is difficult to separate it from other higher functions such as memory, perception or thought. That is why, examining the history of the theories that have tried to explain it we see that the more knowledge that is gathered on the subject, the more complex said theories become. We are now going to introduce some of the theories that have emerged, following the indications of De la Fuente and Pousada (2005).

Selective attention

This historical increase in complexity started with Broadbent theories. The Broadbent Filter Model (1958) was the first attempt to explain how we process and select the most relevant information from the huge amount of sensorial information we receive from our surroundings. This model was the first to introduce the concept of attentional filter which implied the presence of a limited channel. As we can see in the Figure 3 the filter Broadbent suggested was an on/off switch through which the sensorial information either passed or did not. Moreover, this filter was applied in the early stages of sensorial perception, just after the sensory stimuli has passed through the sensorial memory.

Deutsch and Deutsch (1963) approach the same processes in a different way. For them, the selection process was not applied as early as Broadbent theory states. Deutsch and Deutsch suggested a post-categorical filter, creating in that way the Late Selection Theory. This process implied that attended and unattended stimuli are processed on a semantical level and then, according to a set threshold, selected to pass through to the active memory or attention by the filter. Refer to Figure 4 for a visual representation of set filter.

As we have seen, Deutsch and Deutsch‘s model evolves from Broadbent’s early-selection filter to a late-selection filter but the process at the end is still a dichotomous switch. Treisman (1964) argued in his attenuated filter theory that this process was not an all-or-nothing process but a linear one. He stated that the unattended stimuli were not deleted completely from the system but only attenuated.

The concept of a single and limited attention pool underpins all of those studies. In the late 60s, the debate started switching from this basis to models based on resources. This meant that instead of basing the selection and filtering process on the structural characteristics of the system, researchers started thinking that the filtering between different sources of stimulation could be based on the effort that each source required from the available capabilities of the attentional system.

Kahneman model of Attention

Kahneman’s (1973) model (see Figure 5) is based on mental workload. According to him, we have a limited capacity that we can distribute to different sources of stimulation according to necessity. On the one hand, our performance in a certain task will be conditioned by the amount of resources we have available at that time and how we distribute them; and on the other, the characteristics of the stimuli and the amount of attention that the set task requires.

Two of the central aspects of this theory are the allocation policy and the evaluation of demand. These two mechanisms are what determinates the efficiency of the effort is being made. As the author describes it: "The evaluation of demands is the governor system that causes capacity (or effort) to be supplied, as needed by the activities that the allocation policy has selected." (Kahneman, 1973, p. 11).

Regarding this set of resources available Kahneman addresses the concept of arousal. An increment in arousal relates to an increment in the capabilities available to the subject for a certain task. However, this relationship between arousal and performance is not linear. Depending on the task and the context, an optimal arousal level has to be found to ensure optimal performance.

While attempting to find out how the optimal level of activation is found researchers realised that there were two main types of situation that affected the capabilities of our system: the amount of data the system has and the resources it has available to process and respond to the available data.

Resource limited and data limited situations

As we were saying, the resource allocation mechanism response does not fit to a linear pattern when we cross the arousal allocated with the performance obtained. For instance, Kahneman highlights Yerker-Dodson and the effect of noise on task accomplishment as an example. Norman and Bobrow (1975) introduced the concepts of resources-limited and data-limited processes to help understand our capability of accomplishing a task. In a resource limited situation, an increase of the amount of resources dedicated to a certain task leads to an improvement in performance. Consequently, depending on the distribution policy and the correct assessment of the arousal equilibrium point for that specific task our performance therein is going to be affected.

In contrast, when it comes to data-limited situations the chances we have of success are limited by the amount of data we have. Consequently, if we try to impose a variation on the attention distribution, it is not going to change the outcome of our performance, as this is determined by the stimuli itself and the surroundings (data quality).

Nevertheless, Posner and Boies (1971) in their description of the components of attention realized that this limitation on the resources could not only come from the stimuli or the available resources nor could those resources be universally allocated. Based on their research, they stated that when the similarity of the stimuli, either in the way it is presented (same sensorial channel for example) or in the way that is been required to answer to set stimuli are similar, the system gets overloaded easily then when those are different. They concluded that these resources are not universal but specific to the sensorial and motor path through which they are being perceived or reacted to. In the same line, McLeod (1978) carried out experiments using response times to acquire data that backed this hypothesis. He considered that this was an indication of the resources available for the system. He assumed that as the capacity demanded by the task increased, the response times were going to increase.

These two main theories, Paivio's Dual Coding Theory and Kahneman's attentional model are the bases we use to establish a paradigm to understand our experimental situation. When faced with subtitled content we isolated attention as a modulator and the codification of information as a factor in our ability to comprehend the content displayed. In the following section we are going to describe how these two theories connect together.

Dual Coding theory and Kahneman's model of attention

In order to present the theoretical framework of this study we have drawn Figure 6, where we can see a representation of the conjunction of both Dual Coding Theory and Kahneman's attention theory. We can see each channel of information there, [1] audio and [2] visual. Each sensorial path transmits the information separately. This information is processed by the corresponding perception system that has access to a certain amount of [3] Arousal determined by the [4] Resource Allocation Mechanism conditioned by the [9] Demand Evaluation Mechanism.

This has several implications. For instance, the resource allocation policy applied by the Resource Allocation Mechanism and the Demand Evaluation Mechanism are the main determinants of the amount of information the representational systems (the logogens and imagens) are going to receive. This means that the recognition or activation of these basic representations in order to process a stimulus can fail or succeed depending on the bandwidth that the Demand Evaluation Mechanism decides to apply to the Resource Allocation Mechanism.

Going back to Figure 6, the information that passes through the [4] Resource Allocation Mechanism is processed by the Representational System. The representational system is composed of the two subsystems previously mentioned, the [5] logogen and the [6] imagery. The stimuli are processed by the corresponding system: the [6] logogen or [7] imagen. It is important to point out, that at this point the channeled information does not need to follow a sensorially determined path, as can be seen in Figure 6. Once the resource allocation is passed, the information is no longer sensory-specific. Although the imagen system is fundamentally nourished by images, in some situations it can be triggered by words. Accordingly, the logogenic representation system has a visual representation that is written words.

Finally, each set of processed information generates a [8] Response. In Kahneman's attentional theory this represents the result of the processed information and selection of the available actions. But with a broader approach, the response does not need to be the end of the circle but a consequence of the Demand Evaluation Mechanism that will in its turn affect the Resource Allocation Mechanism. If we take the examples of Sadoski and Paivio (2004), in all the DCT, there is an evident link between the sensorial system in the first steps of the process and the motor response it generates because "sensory systems are linked to motor response systems in perception (e.g., eye movements, listening attitudes, active touch)" (Sadoski and Paivio, 2004, p. 3).

Consequently, if we take eye movement as an example, the systems circle around this motor response to adjust its performance to the task or stimuli it is presented with thanks to the [9] Demand Evaluation Mechanism. For instance, if we consider a redundant information scenario where the subject receives information from an auditory channel and through text. This means that the subsequent processes will have to adapt the motor response of the eye to the needs of the representational system. If the system needs more information to come through the channels to be able to recognise the stimuli, the motor response will have to adapt to process more information and meet the demand.

Eye tracking

Eye tracking is a psychophysiological measure that has a long history in several disciplines, from neuroscience to market research. Eye-tracker technology has evolved exponentially in the few last decades, going from EMG's (with no precision over the location of the eye) to infrared light-based equipment. This technological advance has made it possible for a lot of researchers to access the technology and acquire new knowledge in an array of different fields (see Duchowski 2002 for a review).

The debate in many of these disciplines concerns the relationship between eye movement and attention. The visual field of the eye can be divided into three different parts, foveal area (central 2º), the point where our eye has the highest resolution and is able to process more detail; the parafoveal area (5º), where only some stimuli can be perceived or recognised; and the peripheral vision, where the quality of the image does not allow proper perception of the stimuli. These regions of the visual field have applied implications in the perception of stimuli. As described by Sanders (1993) when stimuli are presented in the fovea, no movement is needed to process it. If the stimuli are presented in the parafovea, some eye movement is needed in order to process it. Finally, if stimuli are presented in the peripheral vision, most of the time a head movement will be triggered in order to focus on said stimuli to be able to process it. With this information we can calculate how and where our stimuli are going to be presented in any experimental situation. Depending on the aim of the study, we need to consider these measurements in order to understand the data we are collecting.

There are then two main questions related to the eye movements that Duchowski (2002) raises: [1] in which manner eye movements are controlled to service information acquisition and [2] how we acquire, represent and store information about our visual environment.

Henderson and Hollingworth (1998), argue that the relationship between the aforementioned visual areas and attention is linked to the type of task the participant is performing. In a non-cognitive-demanding task, it is normal to think that the peripheral vision may be gathering information and diverting some amount of attention. Yet when the cognitive load is high, this difference between the focalisation point and the attention decreases. This is due to the fact that in complex tasks, efficiency is crucial to the acquisition of all the visual information needed to process said visual-cognitive task. Rayner (1998) tries to cast some light on this in a review of the studies into reading and eye tracking carried out up to that time. In this review he states “Although we can easily decouple the locus of attention and eye location in simple discrimination tasks, in complex information processing tasks such as reading, the link between the two is probably quite tight” (p. 375). Following the same line of reasoning, Kahneman (1973) states: “the detection of rare events [in the peripheral vision] in the presence of a primary task depends on the allocation of spare capacity to perceptual monitoring, which diminishes when effort and arousal increase” (p. 38).

The second question raised has more to do with the relationship between eye movements and stored information. How a visual pattern can affect the information processed or vice-versa. Entering too deeply into scene processing systems is beyond the scope of this thesis but a general understanding of it is needed to see how it affects our experimental situation.

On this note then, multiple studies (Loftus, 1972; Loftus, 1981; Rayner 1998; Rayner, Smith, Malcolm, & Henderson, 2009) support the notion that our scene processing occurs in two differentiated stages. In the first fixations, a broad overview of the scene and its global information is processed. Once that has been done, we start to process the details of the objects or the portrayed scene in order to glean complementary information.

These image processing strategies have already been proved to help with the processing of the scene in our experimental situation. The aforementioned study by Perego, Del Missier and Mosconi (2010) found that the capability of the participants to recognize scenes and subtitles of the clips they presented was very good and as they say "highlighting the absence of a tradeoff between subtitle recognition and scene recognition" (Perego, Del Missier & Mosconi, 2010, p. 262). Unfortunately, Perego, Del Missier and Mosconi (2010) did not consider in their study the distribution of the time spent between the subtitles and the scene, only reporting that a mere 40% of the participants' fixations were on the scene. This information may seem contradictory to the distribution that has been found to be the norm; however, the subtitle area had more, albeit shorter, fixations than the scene area. Moreover, they only analyze the subtitles on the clip where two lines of text were presented, meaning that there was more information to process compared to what we would have had if we averaged all the subtitles on that clip.

Nevertheless, this presents more controversy related to the degree to which reading subtitles is automatic. In 1987 study Géry d'Ydewalle, Johan Van Rensbergen, and Joris Pollet presented some evidence of automaticity in a chapter called Reading a message when the same message is available auditorily in another language: The case of subtitling published in In Eye Movements: From Physiology to Cognition.

The two experiments presented in this chapter test different variables in the way the subtitles are presented to see their effect on viewers' behavioral patterns. The experiment was based on eye tracking results registered during the viewing of three different clips in three different conditions (audio, no audio and language proficient). The subtitles were also presented in three different time rules (2s, 4s and 6s).

The first important conclusion for us is the time distribution between the scene and the subtitles. According to their results, in a 6s condition (the most commonly used time by broadcasters), where there was no audio, the time spent on each area did not change with respect to the audio condition. Moreover, in the same time situation, participants who were evaluated as proficient in the original language of the audio, spent the same amount of time on the subtitles.

Surprisingly, no differences were found in the sound vs no sound conditions, meaning that participants spent the same amount of time reading the subtitles even if the auditory channel was open and even if they knew the language perfectly. This finding lead d'Ydewalle, Van Rensbergen, and Pollet to conclude that the reading of the subtitles is a process elicited automatically empathizing the fact that "reading a written message is faster and more efficient than listening to the same message" (d'Ydewalle, Géry, Van Rensbergen, & Pollet, 1987, p. 320).

In these same terms, Perego, Del Missier and Mosconi (2010) have established that even though there is an overlapping of information in the visual channel, this overlapping might have been solved by learned processes that allow us to be efficient in this situation and thus not miss any information in the process.

Were it a learned process though, one may wonder how children would behave in the same situation. D'Ydewalle and Bruycker (2007) performed a study to compare the visual pattern of adults and children to find that there are few differences between 2nd grade children and adults. Moreover, they compare two different subtitle presentation options. Standard subtitles (unknown language with known language subtitles) or a reverse subtitles situation (known language with unknown language subtitles). In the past studies such as D'Ydewalle, van Rensbergen, and Pollet (1987) found that contra-intuitively, even in the reverse situation, participants spent 30% of the time focusing on the subtitles. However, D'Ydewalle and Bruycker (2007) found that this attention distribution in their experimental situation was much lower. 41% of the time was spent on subtitles in the standard situation and 26% in the reverse situation. Yet given the situation, this percentage is still higher than expected. The participants spent 26% of the time reading subtitles in a language that they did not understand.

D'Ydewalle and Bruycker theorize as to different possible explanations for this behavior. The most commonly accepted reason is that reading the subtitles has become an automatic process or an Orientation Response situation to the appearance of a new stimuli. However, were this through, D'Ydewalle and Bruycker argue that the time between the subtitle presentation and the first fixation on it would be exactly the same in both situations (standard and reversed subtitles) yet it is higher in the reverse situation (333ms vs 416ms). As Kahneman (1973) explains, the OR is a response that decreases easily with habituation within a learning process. In our situation, when we talk about automatic reading, this needs to be differentiated from an OR response. We can not expect a Orientation Response in subtitles reading if (as is probably the case) the participant has been exposed to subtitles before. It is most probably this lack of differentiation what makes D'Ydewalle and Bruycker reach their conclusion. Altogether the experimental situation they generated was ecological (no instructions were given about what to process or what to do during the video presentation) it lacked some basic verifications. Given that there were neither comprehension questions nor visual recall questions, the participants were free to behave as they saw fit. One of the things that we need to test is what happens if we lessen this freedom with questionnaire verification that forces participants to process both, textual and scenic information. A possible explanation of the lower yet significant amount of attention dedicated to reverse subtitles might be related to an arousal surplus in the visual channel. As the authors say, the selected scene was one with a lot of dialogue, we also assume that given the characteristics of the cartoon presented ("Pelle Svanslös"), the scene was somewhat simpler and slower in comparison with a typical film. This would have meant that the amount of arousal available for the visual channel was greater than the requirements of the scene causing the drift towards a new, changing stimuli, the subtitles.

Given then that most of the studies seem to argue that the percentage of time spent on each visual area does not change, Bisson, Van Heuven, Conklin and Tunney (2014) raise an interesting finding concerning differences in fixation time and fixation count. He examined differences in subtitle reading among English native speakers in three languages and in the most common subtitle conditions (interlingual, reversed subtitles and intralingual). Having analysed the number of fixations, mean fixation durations and the number of subtitles skipped, they found that although people read the subtitles in all the conditions, a more regular reading pattern was observed in the case of standard Dutch-to-English subtitling where viewers did not know the language of the soundtrack, but knew the language of subtitles.

As expected, participant's fixation durations were longer on subtitles when they were in a foreign language (either with Dutch or English subtitles) than when they knew the language.

All the control group comparisons allowed them to make sure that those fixations in the subtitles area were not due to film action happening there. This use of the control group seems questionable at the very least. If a good control of the area of interest and the subtitles presentation was done, this comparison is pointless.

Another error they committed was to trying to normalise subtitle presentation between the two languages by using the number of words. As they themselves say, word count was not the same in the English subtitles condition as in the Dutch condition. Following the same logic, of course letter count was not the same either, consequently, a normalisation based on the character count would have been more accurate.

Objectives

- To establish if there is a relationship between the audio language and the subtitle reading process and analyse how this relationship affects comprehension. The allocation of resources is a dynamic process. We want to assess whether or not the allocation of resources changes depending on the language of the audio and the effect this has on the processing of the content.

- To establish the attention distribution between the visual and textual content, how this varies across different devices and how it affects the comprehension and recall of the visual and textual content.

First Experiment

Divided Attention and Redundant Information Channeling in an Ecological Situation

Introduction

The goal of the first experiment was to compare subtitle processing in two different situations: when the language of the film is known by the viewers and when it is not. We carried out this experiment in Poland where we wanted to study a typical subtitle reading situation, i.e. people watching a video in English with standard interlingual subtitles in Polish, and compare it with a condition in which the language of the video is unknown to them. We therefore presented our subjects with foreign-language clips subtitles in their native Polish.

For this study we took a holistic approach based on the Dual Coding Theory and the contributions on attention presented in the previous chapter (page 37). The objective of this experiment is to obtain overall image of the process taking the different factors that may be overlapping into consideration. In doing so, we wanted to have valid information on the simultaneous processing of the different signals in different modalities.

Subtitle processing has been studied from a psychological perspective and from disciplines such as translation and subtitle generation.

On the one hand, we are going to take the psychological perspective of the task into consideration. From a psychological point of view, as mentioned before, the main factor to consider is the attentional resource distribution. Historically, the few studies conducted by psychologists on subtitling have used this specific situation as a task to understand the nature of this resource allocation procedure, which is occurs automatically in our attention system (see Perego, Del Missier & Mosconi, 2010; d'Ydewalle, Van Rensbergen & Pollet, 1987). On the other hand, translation and subtitle creation studies have focused on the accuracy of the presentation of the subtitles and the language learning aid that they may provide (see Kruger et al. 2014; Mayer & Moreno, 1998; Paas, Tuovinen, Tabbers & Van Gerven, 2003).

From our point of view, one approach may not be understood without the other. The attention allocation is responsible for the resource allocation to the different channels involved in the coding or the dual coding of information, serving as a modulator for any effect that this may or may not have.

As seen in the introduction, two main psychological theories underpin this experiment. On the one hand, DCT predicts a better comprehension of the content in a situation of redundant information. In our experimental situation, the redundancy appears in the English clips: Participants understand the audio of the clips and can read the subtitles, thus having two different (and redundant) sources to complement the information transmitted. On the other hand, Kahneman's (1973) attention theory provides a key concept: the resource regulator, responsible for managing the amount of information delivered to the cognitive system.

Mayer and Moreno's (1998) experiment is a good example of how these systems work together. They presented the participants with two different experimental situations. In one, a group of students had to watch an animation on how lightning formed with a concurrent auditory explanation. In the second one, they had to watch the same animation but with an explanation in text on screen. In all the comprehension questionnaires completed by participants, the group with auditory narration outperformed the one with text explanations. As they theorised, "given the limited resources the students have for visual information processing, using a visual modality to present both pictorial and verbal information can create an overload situation for the learner" (Mayer and Moreno, 1998, p. 313).

This gives us an idea of how important the Demand Evaluation Mechanism is in information processing. According to Kahneman (1973) The Demand Evaluation Mechanism is defined as "the governor system that causes capacity (or effort) to be supplied" (p. 11). That said, it is important to point out that the Resource Allocation Policy is not only affected by the Demand Evaluation Mechanism. Kahneman states that three more factors determine the Resource Allocation Policy: the enduring dispositions, meaning involuntary attention (e.g. orientation response); the momentary intentions (e.g. listening to a specific voice or music); and systematic effects that may occur when we reach saturation.

In our experimental situation, the interaction of these two mechanisms leaves us with a stimulus that carries three types of information: (1) The subtitles processed by the visual channel and corresponding to the logogenic representation of information. (2) The scene processed by the visual system and corresponding to the imagery representation system. (3) The audio, processed by the auditory channel and corresponding to the logogenic representation system. All these factors are interconnected, as Duffy (1992) states, even if we consider them to be two independent systems, connections between them must be assumed. In a picture-text situation like the one we face with subtitles, both systems work together towards an "integrated representation of the message" (Duffy, 1992, p. 463).

When the saturation is related to the coding channel, in this situation the visual or auditory system, the information received and processed by the two background systems (imagery and logogen) is affected. Having said that, hitting the logogen system with information perceived by multiple channels enriches the input that the system has thus improving the response, as long as the information on both channels is coherent. Therefore, we have a primary system that has a set amount of resources and whose ability to manage multiple hits is limited, whereas the background system can manage and in fact merges and takes advantage of multiple channel input.

Nevertheless, this multiplicity of sources needs to be separated from the noise. The attentional system has to pay the right amount of attention to each channel and select the relevant information from it. The degree to which the attentional system can benefit from this redundancy (and indeed identify it as a relevant factor and not just noise) depends on a number of factors such as proficiency of the language, volume, reading skills, etc. Viewers proficient in the foreign language will benefit more from redundancy, as they will be able to complement the information presented in the subtitles with that coming from other channels. In contrast, viewers who do not understand the foreign film dialogue will tend to classify the information that comes from the auditory channel as noise and as a consequence, they will be less likely to draw benefits from the verbal auditory channel, as this kind of input will not be comprehensible and will thus remain unprocessed.

As an example of the correlation between these two systems, Duffy (1992) draws a parallel between the lexical recovery system when we read or process language stimuli. As we know, the lexis we have stored activates in relation to the context of the stimuli we are reading thereby facilitating access to lexical items that are more closely semantically related to the text we are reading. What Duffy (1992) theorizes is that one of the possible intersections where imagen and logogen information merge are in the facilitation of the lexic in a similar way to when we only have text based stimuli. Although the verification of this possibility is beyond the scope of this thesis, the example serves to understand how both systems can relate to one another.

Rationale and hypotheses

In concordance with the information retrieved from different theories, in our study better comprehension and scene memory will be found in the English version of the clips. The logogen informational system will receive two different forms of input. This redundancy will be shown in the results of the comprehension questions, which will be better in the English-Polish situation than the Norwegian-Polish situation.

At the same time, this duplicity of information is going to affect the demand evaluation mechanism of our attentional system. This is going to translate into a behavioural change: a difference between the percentage of time fixated on the subtitle area vs the percentage of time fixated in the scene area will be found; when the audio is in Norwegian (unknown language) more time will be spent focusing on the subtitle area. When the audio is in English less time will be spent in the subtitle area. As this situation will be perceived by the system as easier than the mono-input situation, the system is going to adapt accordingly.

Finally, as the presentation of the subtitles is exactly the same in the different situations, no difference is going to be found in the fixation count and the fixation duration.

Method

Design

In this within-subject study, each participant was shown two videos from a set of three, one in each language. We varied the order of presentation for the different participants. After each clip, the participant had to answer 10, multiple choice, textual and visual recall questions: five concerning the information presented in the text of the subtitles only and five on the visual aspects of the video. Eye movements were registered during the viewing of the clips.

Participants

The participants were eight men and twenty-five women, whose age ranged from 20 to 35 years old. The mean age of participants was 24 years old (s=3.4). To assess their subtitle viewing habits we asked how often they watched foreign films with subtitles, over half of the participants declared that they sometimes do (54%). Many of them stated that they regularly engaged in this activity (39%) and only six percent of the participants admitted that they do it rarely. For the language proficiency evaluation, we used self reported measures. In a similar way to Bisson, Van Heuven, Conklin and Tunney (2014), we used a self-reported language proficiency questionnaire to ascertain the participants' proficiency in English and Norwegian. Participants were asked to mark their declared proficiency on one to ten scale (one – no knowledge, ten – proficiency). The declared command of English was much higher than that of Norwegian. The mean proficiency in English was 8.79 (s^2=1.19) and 1.36 (s^2=.55) for the Norwegian.

Materials

SMI RED eye tracking system with a sampling rate of 250Hz was used to record the participants' eye movements while they watched the two sets of videos. The experiment was prepared using SMI Experiment Center. This eye tracker has a 22 inch monitor with a resolution of 1920 by 1200 pixels.

SMI BeGaze software was used for fixation and saccade detection. Minimum fixation duration detection algorithm was set to 80ms. For statistical analysis and data preparation we used SPSS.

In the analysis, areas of interest (AOI) were drawn separately on each subtitle with frame-accuracy and the rest of the screen area was considered scene (noAOI).

Stimuli

The video fragments used in the study were taken from the English film Joining the Dots, directed by Pablo Romero Fresco (2012), and the Norwegian film Headhunters, directed by Morten Tyldum (2011). Each video fragment lasted between 3 and 4 minutes. The fragments were cut so that they contained an entire meaningful scene.

The subtitles were prepared in EZTitles subtitling software, using the following settings: maximum number of characters per line 38; the maximum number of lines was two, the second larger than the first aligned in the center; the subtitle presentation rate was 15 characters per second; the maximum exposure time was six seconds with a minimum exposure time of one second; the gap between the subtitles was always more than three frames; the font was Arial displayed in white. The total number of subtitles analyzed in this study was 295 (175 in English and 120 in Norwegian).

Questionnaire

To evaluate the different aspects involved in the processing of the video, we prepared a 10 question questionnaire with two different sections. 5 questions were related to the textual content of the clip and 5 to the visual content.

The textual questions of the questionnaire were based on the content of the subtitles. For the visual questions, a previous study was conducted to establish the salient aspects of the clip. 32 participants saw the clips and after each one they wrote down five objects or visual clues that they remembered. The questionnaire was developed based on the frequency with which each object appeared. We followed this methodology based on the procedure used by Lavaur., and Bairstow, (2011).

Procedure

All participants were tested individually. After being welcomed by the researcher, they signed the informed consent form and answered the demographic questionnaire. They were then assigned to a random order of clips. After that they were seated comfortably at a distance of approximately 60 cm from the monitor.

After the standard 9-point calibration and validation procedure, participants watched the video clips in counterbalanced order according to their designated order. They were instructed to watch the clips carefully, as they would have to answer some questions related to the clip content. After each video, they were asked to complete a 10-question textual and visual comprehension questionnaire displayed on the monitor. The whole procedure lasted about 30 minutes.

Results

In the first stage of the analysis we eliminated the variability factor of the between the clips. In order to do that, we had to control the differences in length, number of subtitles and number of characters in those subtitles. To do so, we took the eye tracker variables and standardize them with the relevant variable to control. In that way, we calculated the number of characters per fixation (dividing the number of characters in each subtitle with the number of fixations on that subtitle), normalized dwell time (time that the subtitle was on the screen divided by the time spent in the subtitle area) and finally, the visit count standardization (number of subtitle in the clip divided by the number of visits on the subtitle area).

Once all those corrections were done we proceeded to the main analysis. Two repeated measures ANOVAs were performed grouping the data in two different meaningful groups. The first group contained all the metrics and data concerning the textual content of the subtitles: characters per fixation, average fixation duration and the results of the textual questions of the questionnaire. On the other hand, the other ANOVA grouped the data related to the visual content. he dwell time was also included, along with the standardized visit count and the results of the visual content questionnaire. For all the eye tracker measurements, we considered two areas, the subtitle area (AOI) and the scene area (noAOI).

Textual content results

3 of the participants had to be excluded from the analysis due to either outlier scores in at least one of the eye tracking measurements or to trackability problems.

Regarding textual content data, the repeated ANOVA measurements showed differences between the languages in the different dependent variables ($F_{(3, 27)}$=8.67, p<.000). If we take a look at the univariate contrasts, we can see where the differences are located. The results, as presented in Graph 3 show that the main differences are in the Number of Characters per Fixation ($F_{(1, 29)}$=20.21, p<.000, w^2=0.26) the Norwegian characters per fixation (x=6.43, s^2=1.77) was significantly higher than the English average (x=5.10, s^2=.73). However, the Average Fixation Duration does not change significantly between languages ($F_{(1,29)}$=.052, p=.821, w^2=.01) see Graph 4. Finally, we can also see significant differences in the Questionnaire results ($F_{(1, 29)}$=4.337, p=.046, w^2=.08): the scores for the English clip (x=4.40, s^2=.814) were significantly better than the scores for the Norwegian clip (x=4.03, s^2=.809) see Graph 5.

Visual Content results

4 of the participants had to be excluded from the analysis due to either outlier scores in at least one of the eye tracking measurements or to trackability problems.

We performed the same repeated measures ANOVA on the visual content data. The results show that the language has no major effect on the variations shown in the variables ($F_{(3, 26)}$=1.87, p=.157). This means that the language of the audio played does not affect the distribution of the attention between the subtitle area and the main scene area (AOI and no AOI) this is illustrated by the fact that there were no differences in the responses to the questionnaire or in the eye tracking metrics registered. See Graph 6 and Graph 7.

Discussion

The aim of this study was to examine differences in the processing of subtitles depending on the original language of the video soundtrack. The main research question we set out to address was whether or not proficiency in a foreign language affects the comprehension and viewing patterns of Polish viewers watching English and Norwegian videos with Polish interlingual subtitles.

The study produced a number of interesting findings. First and foremost, in answer to our main research questions, we found statistically significant differences in the way viewers process subtitles depending on the audio language of the video, but not in how they process the visual information on the scene. As we can see in the results, there is a clear variation in the way we process subtitles depending on the audio input we are receiving. Yet this does not interfere in any way with the processing of the main image of the video content. This is illustrated by the fact that we did not find any differences at all in the scene content, in the textual content we found differences in characters per fixation and in the questionnaire performance.

Moreover, this study further confirmed that people read subtitles even if they understand the language of the video - an effect found previously by d'Ydewalle and

Gielen (2007) and aptly referred to by Duffy as the “magnet effect of subtitles” (1992, p. 464) -. In this case this effect was particularly visible with English videos watched by participants who were proficient in this language but not in the way we hypothesized. Instead of finding a difference in time distribution between subtitle area and main scene are, the difference was found in how the subtitles themselves are processed. Consequently, proficiency in the language of the original video was linked with higher comprehension scores, in line with our expectations. As predicted by the DCT, information processed both visually and orally aids comprehension significantly (Paivio, 1986).

As we have shown in the results, the data we gathered was divided into information obtained on the processing of the textual information and information on the visual content of the scene. This separation led to us finding some interesting differences that we would not have found without it.

On the one hand, the differences in the soundtrack of the video did not affect the time distribution between the main scene and the subtitle area. Given this lack of difference in our visual performance, it seems only logical that the performance in the processing of visual information in both situations would remain the same. The ability to process the visual aspects of the video remain unaffected by the language of the audio to which we are listening. This non-significant result could be a result of the process elicited by the experimental situation similar to those elicited in Kruger, Esté and Matthew (2014) given that, like the students in the classroom, the participants in this study needed to learn and pay close attention to the detail to be able to answer the questionnaire.

On the other hand, where we were able to find differences was in the processing of the textual content. First, we can already see that there is a significant difference in the performance of the participants in the answers they gave to the questionnaire. As we were expecting, the results of the English questionnaire are better than the those of the Norwegian one. These results support the idea that greater language proficiency has a positive effect on comprehension but the question of the processing of the content remains unexplained.

This provisional confirmation of the hypothesis contrasts with the results shown by the eye tracking data. Even though we were expecting the results to show a faster reading pattern for the English clips due to greater knowledge of the language, the results do not support this. The data shows that the participants' performance (characters read for each fixation) was better in Norwegian, meaning that they were reading the Polish subtitles faster in the Norwegian clips.

This situation can be explained if we go back to the dual processing theory and its relation with Kahneman's theory on demand evaluation. In the Norwegian film they only have one source of information, the subtitles, as they do not understand anything from the audio channel. This generates a necessity to read all the information displayed in each subtitle, as they cannot complement the information in any other way, any lost time implies a loss of information. This need puts pressure on the participant and leads to a decrease in the quality of the information they are able to process and retrieve, hence the lower scores in the questionnaire.

However, when the soundtrack is in a language they know, the characters per fixations drop. The pressure to gather information as quickly as possible disappears, as they know that they can complement the information they are getting through the visual channel with the audio channel. So we could say that in this situation, the participants are more concerned with the quality of the information they are gleaning than in the quantity of it.

It is important to notice that the subtitles still appear to be the main source of information in this instance. That is probably because if the retrieval of information was switched completely from the subtitles to the audio, we would have found differences in the visual content results as well. Besides, the lack of variation in the time distribution between the subtitle area and the scene, seems to be in line with this explanation. However, we tend to assume that in this situation the main information withdrawal would have been carried out by the auditory system. If we look at the data in more detail, we find some differences in fixations that may be an indicator of this. While the fixation duration did not vary, the number of fixations was significantly different, in relation to the number of characters.

Mean fixation duration is often taken as an indication of processing difficulty. Under this premise we would assume then that the fixations in the Norwegian clip would be longer. The lack of difference could indicate that the English clips did have some extra processing load. It is possible that when watching the videos in a language they knew, the viewers tried to compare the English audio with the Polish translation in the subtitles, which in turn resulted in greater effort and longer processing. When watching Norwegian videos, viewers' reading resembled regular reading but, given the channel saturation (which we will explain later) and the lack of auditory aid, the load might have been the same in the end.

Going back to the explanatory theories, Figure 7 helps us understand the processes taking place in this situation. First we have the stimuli characteristics divided by channel: [1] the language audio (English or Norwegian) corresponding to the Auditory Channel; [2] the interlingual subtitles and [3] the scene, both corresponding to the visual channel. In the Norwegian situation, the audio is only processed marginally, while in the English situation, the path follows towards the [6] logogenic representational system. On the other hand, subtitles and scene are both processed by the visual system. The information corresponding to the subtitles, is diverted towards the logogenic system. However, the scene information follows its path to the imagen representational system.

In this theoretical representation, in the English clips we have a situation where the logogenic representational system has two sources of information coming from two different sensory pathways, whereas, in the Norwegian clips, each representational system has only one source of information coming from one pathway. As we have found in the results, in this last situation, the pathway gets saturated by the amount of information it needs to accomplish the task. The amount of Arousal that the [10]

Demand Evaluation mechanism can provide to the pathway is not enough for it to succeed.

Interestingly, in the other experimental situation, this saturation is not reached. We can see that the Demand Evaluation mechanism has no problem providing the amount of arousal needed to succeed in the task. One pathway helps relieve stress on the other pathway. Now though, instead of having a single source of information for each representational system, we have two forms of input on the logogenic representational system. Theoretically, this could also represent a saturation of information for the system but at a higher level. However, the data we have analysed supports the DCT and demonstrates that this multiple input situation is correctly processed by the system as redundant information. As a consequence, participants' responses to the questionnaire improve.

As a direct result of this last logical reasoning, one last factor needs to be addressed. We have already seen that we can saturate the pathway from which the system receives information. The remaining question would be to see if we could saturate the representational system using translation errors, for instance. Often when we are watching a subtitled film we see a word in the subtitles that does not correspond to the actual spoken word. Further research could address this situation to see if the representational system has a limited capacity, as we have seen with the sensorial pathway.

In conclusion, when it comes to Dual Coding Theory and the attention allocation system one cannot be understood without the other in this situation.The ease and efficiency provided by the ability of our perception system to process multiple, synchronized sources of information has proven to be key to attention allocation and information processing. Further research using subtitles must be done to further explore the interaction between these systems.

Limitations of the study

Different video genres (documentary and fiction) were used for each language. Although we did corrections and standardizations to tackle the different characteristics of the video that may affect the results, we still need to determine if the genre generates significant differences in this kind of processing.

There is a desperate need for standardization on the subtitle studies. To be able to compare results, some standards need to be established. The calculation of character per fixations, the percentage of subtitles skipped or the actual presentation of the subtitles themselves need to be standardised in order to understand how different studies relate to one another.

Second Experiment

Attention distribution in content visualization in different devices

Introduction

As we pointed out in the general introduction (page 19), the technological revolution has had a significant impact on the way that younger generations consume content. Mobile devices are becoming more important due to improvements in their efficiency and the fact that they are portable.

Moreover, the way in which viewers access content has also undergone something of a revolution. Youtube, Netflix, Amazon Prime, etc. are only a few examples of platforms that are natively developed and compatible with mobile devices. As a result, they are promoting the use of such devices to consume their specific content, thus making them (in some cases) the preferred method of content consumption.

Data from Zenith's Media Forecast (2016) offers a fascinating insight into this trend. They found that people spend on average 19.7 minutes a day consuming media on mobile devices compared to 16 minutes a day on fixed platforms such as smart TVs. In their forecast, they predict that use of mobile devices will grow to the point where they will be the main platform for media consumption, with people spending 68% of their time on them, by 2018.

But could the platform change the way we process information? All this evolution at the same time may affect the way we process the information being transmitted to us. We need to assess the effect that different screen sizes can have on the information recollection, in other words,can these differences affect our information processing and memory? e also need to understand if a different approach to subtitles is needed to improve information recollection on those different platforms.

In order to test this situation we carried out a second experiment in Spain in parallel with our Polish experiment. In this second experiment, we tested two different devices, a tablet and a mobile phone and compared their performance and the visual strategy used by the viewer to the normal desktop setup.

We used eye tracking to study users' visual exploration when watching clips on these three devices, plus two questionnaires, similar to those used in experiment 1, to evaluate visual and textual content comprehension and memory.

Any differences in the attention allocation in conjunction with differences in the ability of the participants to answer visual and textual questions could indicate problems with the processing of the signal. In this experimental situation the screen size may affect the resource distribution policy of our attentional system thus affecting the balance of resources dedicated to each signal of the stimuli (subtitles and scene area). These modifications of the resource allocation, can affect the amount of information the system receives in turn affecting comprehension of said content.

If we go back to our theoretical schema (see Figure 6) of the system. In this experimental situation we eliminate the auditory channel by making it incomprehensible for the participants (clips in Norwegian). This meant that the only source of information was mono channel thus disabling any facilitation effect produced by information redundancy. In this situation, the Resource Allocation Mechanism had to manage different activation through the same system directed at two different sources of information.

Hypothesis

The responses of the questionnaires (one for the visual information and another for the textual information) will show a gradual improvement in relation to the screen size as a bigger screen will act as a facilitator to the processing of the content.

The distribution of the total time spent between the subtitles and the scene area will change depending on the device. On the Smartphone the participants will spend more time on the subtitles and this will decrease as the screen size increases.

Methodology

Design

We used a within-subject design. Each participant watched one different clip on each of the devices (smartphone, tablet and monitor). We randomised the device and clip order following a Latin Square distribution. After each clip, we asked the participant to answer 10 questions about the clip content. 5 questions related to the textual content of the clip's subtitles and 5 questions related to the visual content of the clip. During the three viewings, eye movement was recorded.

Participants

We collected data from 22 participants 11 male and 11 female. Ages ranged from 25 to 35 years old (m=29.23; s^2= 3.74). 55 percent of them always consumed subtitled content, 22 percent viewed it regularly and the remaining 23 percent rarely or never did so. Following the same procedure as Bisson, Van Heuven, Conklin and Tunney (2014) we used a self-reported language proficiency scale to report language proficiency in Norwegian. None of them spoke or understood Norwegian.

Materials

We collected Eye Tracking data using the Tobii X2-60 eye tracker. It has a sampling rate of 60Hz, a precision of .4º and an accuracy of .2º (Tobii Technology AB. 2013). For the devices presentation, a Tobii Pro Mobile Device Stand 2 was used. The software to register and post-process the images and data was the Tobii Pro Studio Professional Edition. The standard I-VT Fixation Filter was applied to classify fixations and saccades (see Olsen 2012 for a detailed description).

The stimuli were presented on three different devices. [1] An Apple iPhone 4 with a 3.5 inch screen and a resolution of 640 pixels by 960 pixels. [2] An Apple iPad Air with a 9.7 inch screen and a resolution of 2048 pixels by 1536 pixels. [3] A 22 inch Acer monitor connected through an HDMI port to an early 2013 MacBook Pro Retina running Windows 7 and Tobii Studio Professional 3.4.4. To acquire data from the presentation of the clips on the mobile devices we mounted the X2-60 in the Tobii Device Stand. This stand holds a webcam that records the content displayed on the mobile device and then projects the Eye Tracker data over this recording. For the monitor, we used the same Eye Tracker device mounted to an Eye Tracker holder placed in the bottom frame of the screen.

The clips were played with VLC media player, on the mobile version and the desktop. The subtitles were configured in the same way on the different devices.

For the analysis, we drew an area of interest (AOI) on the subtitle area and considered the rest of the screen (noAOI) as the scene. This procedure was carried out for each participant and on each device.

Stimuli

Three two-minute clips were cut from a Norwegian film called Head Hunters. The clips were selected so that they contained an entire meaningful scene. The number of words was controlled and the locations where the scenes depicted in the clips took place were different. The quality of the clips was reduced to 720p HD so that the iPhone 4 could play the clips without any issues.

The Norwegian - Spanish interlingual subtitles followed general industry standards. The maximum exposure time was six seconds, the minimum one second. The gap between subtitles was always more than three frames. The assumed reading speed was twelve characters per second. The maximum numbers of lines were two, the second longer than the first. The maximum line length was between thirteen-six and thirteen-nine characters giving a maximum number of characters in total of seventeen-two. The font was Arial, the color was white and the font size was kept proportional to the screen size. To generate these subtitles we used EZTitles. The total number of subtitles analyzed was 120.

Questionnaire

We developed two questionnaires based on the content of the clips in order to test participants' comprehension. Each one of them had ten questions, five on the verbal content (subtitle content) and the other half on the visual content (objects in the scene).

To generate the visual content questionnaire, we used the same method as Lavaur and Bairstow (2011). We presented the clips to 32 students in a classroom and asked them to write down five visual elements they remembered from each clip. The results were analyzed and the questions were written based on the frequency of the objects that were remembered.

The questions for the textual information where based on the content of the subtitles.

Procedure

All the participants were tested individually. The participants were assigned a random combination of devices and clips. The researcher welcomed them and after reading and signing the informed consent form they completed the demographic questionnaire.

The participant was then asked to sit in front of the corresponding device and the 5-point calibration process was started. This process was repeated for each device. Then the participant watched the corresponding clip, they were then asked to answer the corresponding comprehension and visual recall questions. After watching all the clips on a different device each time, the participant was thanked and dismissed. The process took approximately 35 minutes.

Results

We performed a within-subject MANOVA for the questionnaire results establishing the devices as factors and the two questionnaires as variants. For the eye tracking data, we performed two different tests. [1] A within-subject ANOVA: with the devices and the location (AOI and noAOI) as factors and the Total Visit Duration as a variable. [2] A within-subject MANOVA: with the devices and location as factors and the Fixation Count and Fixation Duration as variants.

Questionnaire

The results of the questionnaire show that those viewing the clips on the tablet scored higher, with a correct answer rate of 77.36%. Followed by those who used the monitor with 73.24%. Those using the Smartphone had an answer rate of 67.61%. If we divide the results into comprehension questions and visual recall questions, the results follow the same trend (see Graph 8). However, this distribution does not represent a significant difference between the different devices ($F_{(1,21)}$=1.87; p=.163)

Eye Tracker data

To analyze the eye tracking data, we used a within-subject MANOVA. The first parameter we took into consideration was the total time spent in the subtitle area vs the rest of the scene (Total Visit Duration in Tobii Studio). The results showed no difference between the devices ($F_{(1,21)}=1.01$; $p=.391$). The distribution between the two zones was on average 31.98% of the time spent in the subtitle area and 68.02% spent in the scene area. In Graph 9 we can see a detailed representation of the distribution on each device.

We found some differences while analyzing how participants eye movement varied from device to device. The main effect for the fixation characteristics analysis shows a significant difference between the devices ($F_{(1,21)}=18.47$; $p<.005$) and between the area ($F_{(1,21)}=35.01$; $p<.005$). If we take a closer look at the data, we can see that these differences concern the fixations, which seem to distribute differently in relation to the device. While with the tablet and the monitor the number of fixations does not differ (see Table 1), with the mobile phone the Fixation Count shows a tendency to decrease ($F_{(1,21)}=1.19$; $p=.079$). On the other hand, we also registered an increase in the duration of those fixations. See Table 2 for a summary of the Average Fixation Duration on the different areas and devices. While on the monitor and the tablet the duration does not vary, the fixations on the mobile phone are longer ($F_{(1,21)}=13.78$; $p=.002$).

Discussion

Our aim in this experiment was to determine if different platforms elicited differences in the way we process content. As we explained in the introduction, new viewing platforms are a growing market that present some challenges. In this context and under the scope of the HBB4ALL project, this was one of the first tests that has helped us understand whether or not it is necessary to modify subtitle presentation on mobile devices.

As a general conclusion to that question, we can now say that the differences in information processing on different devices are minimal. Regarding the data collected, although we found some interesting patterns that we are going to discuss, the general conclusion is that keeping the presentation of the subtitles proportional to the screen size is a good approach. It would seem that no more changes to subtitles are required in order to obtain the same levels of comprehension of the content.

Turning first to the questionnaire results first (Graph 8), as hypothesized, we can only see a trend in the answers where there is a relationship between screen size and performance. These differences may not however, be classified as significant. This applies to the visual questions as well as to the textual questions.

Moving on to the eye tracking data, the distribution of time spent between the subtitle area and the scene does not change significantly in relation to the device, maintaining the trend seen in experiment 1. Moreover, this distribution seems to be very similar to that found in the first experiment (see Table 3 for a summary of the time distribution on all the devices). As we can see, on average this distribution is between 30-35% on the subtitles versus 65-70% on the scene.

Looking more closely at the rest of the eye tracking data we start to find differences. It seems that the way the attentional system solves the problem of a narrow screen is by modifying the way information is collected. In this experimental situation fewer fixations in relation to the decrease in screen size is understandable given the fact that there is actually less physical space to cover. To compensate for the increase in information density, this decrease in the number of fixations is balanced by longer fixation times.

It is when we put the questionnaire data and the eye tracking data together, that we find interesting information. The fact that the answers to the questionnaire do not change but the motor response does is in fact very interesting. In these different situations, the information processed is almost the same across all the devices. The responses to the questionnaire, do not show significant differences (although they show a trend) meaning that, as far as we can tell, screen size does not affect comprehension. That said, if the motor response varies but the comprehension does not change, it means that our attentional system and in general, our information processing system is able to adapt correctly to the situation.

Going back to Figure 6, in this situation we face a mono-channeled information-processing scenario. We have the unknown audio in the audio channel and the subtitles and scene in the visual channel. In our experimental situation, the content is watched in three different screen sizes. With some calculations, we can understand what these different changes represent to our visual system. The screen size for the iPhone 4 is 5.8cm in width by 10.4cm in height hence an area of 60.32 cm^2. The screen for the iPad 2 is 14.78 cm in width and 19.71 cm in height giving an area of 291.31 cm^2. Finally, a 22' monitor has a screen width of 27 cm and a height of 49 cm and an area of 1,323 cm^2.

As stated by Rayner (1998) the foveal area of the eye is about 2º while the parafoveal area occupies 5º. At an approximate distance of 60 cm, the foveal area of our field of vision would be approximately 2.1 cm in diameter, that gives us an area of 3.96 cm^2. The parafoveal area is 5.24 cm in diameter covering an area of 21.57 cm^2. That means that in one fixation we cover 6.56% of the screen size of the iPhone, 1.35% of the iPad and .3% of the monitor. The screens were positioned horizontally, meaning that in 4.95 fixations we can scan from one side of the iPhone screen to the other, taking in the entire screen. For the iPad, we need 9.39 fixations and for the monitor, 23.33 fixations.

These calculations help us put the data we have into perspective. Even though there is a substantial difference in the screen sizes, the number of fixations used to scan the subtitles did not change significantly in our results (mobile 189.07 fixations, tablet 203.27 and monitor 201.6, see Table 1). This means that in relation to the screen size that the visual system had to cover, the mobile phone needed a considerably higher number of fixations compared to the monitor screen. As there is no established way of standardising the relationship between space and fixations, the data analysis had to be done with non-relative numbers that, as we can see, are masking the effect.

If we approach the same subject from another perspective, we could say that the fact that it took the same number of fixations to read the content on one device compared to the rest may indicate that fixations are linked to content. As the content between devices does not vary from participant to participant, the fixation count shows different individual variations but it remains the same. If we read the data from this perspective, fixation duration is the factor that comes into play. If fixation count is a content related metric, the increase in fixation duration in our experiment is what gives us information about the experimental effect.

This discussion is a well-known debate when studying reading behaviours. The question is whether saccades are executed to cover the distance between a certain number of characters or to cover a certain visual angle. To address this question, Morrison and Rayner (1981) conducted a very simple experiment in which they modified the reading distance keeping text size constant. Their results showed that the saccadic length of the participants decreased in line with the increase in reading distance, meaning that the amount of characters that each saccade covered stayed constant. The amount of characters covered by saccade in average was 5.43, 5.30 and 5.70 at 36, 53 and 71cm.

Regarding fixation duration, our results also replicate those of Morrison and Rayner (1981). They state that as letters get smaller, the amount of retina space covered by the letters also shrinks making discrimination more difficult, hence the increase in fixation duration. The question is: to what processing level can we attribute the factor that provoked this increase? On the one hand, as we described, we could be facing a mere physiological adaptation. When a decrease in the resolution of the information occurs, the fixation time compensates for the loss of input, as described by Morrison and Rayner (1981). On the other hand, this could also be related to higher processes. As supported by the literature, fixation duration has been linked to mental effort (Just & Carpenter, 1980).

From our point of view, both explanations at are two ends of a continuous. When external demands challenge our physiological ability to respond, the cognitive effort of the task increases. This is the same situation as in the classic experiment conducted by McLeod (1978) mentioned in the introduction. In his experiment, he took response time as an indicator of the amount of resources available to the system. This physiological motor translation from our attentional system to the motor response is what Kahneman's model predicts. Once our attentional system is faced with a demanding situation, our resource distribution, guided by the demand evaluation mechanism, allocates the resources necessary to face that situation. In our experimental paradigm, this translates to longer fixation duration, in a similar way to how, in McLeod's experiment, it lead to an increase in the response time.

In summary, now we can offer an explanation for the first answer we gave to our main research questions. We established that there was no difference between the devices hence no adaptation of the subtitles was needed. The conclusion does not vary as the questionnaire results do not show a decrease in comprehension. We can now say that even if there seems to be no significant and meaningful difference, there is a difference in processing that needs to be explored further. The fact that these differences did not appear in the questionnaire, make the results of greater interest. The mechanism that came into play in our experimental situation managed to compensate for the increased effort needed to accomplish the task successfully (or at least as successfully as it did on the other devices).

General Discussion

A broad image of attention distribution has been presented in this thesis.

We have carried out two experiments using subtitling as an experimental paradigm in order to support an integrative model of information coding and retrieval and its relation to attention allocation, based on Paivio (1986) and Kahneman's (1973) proposals.

Several perception strategies have emerged as a result of the different experimental manipulations performed in the two experiments we prepared in the HBB4ALL framework. In one experiment, participants watched two clips subtitled in their mother tongue; One of the clips was in a foreign language that they could understand and the other clip was in a language that they did not understand. The second experiment evaluated the processing of information while watching clips on different devices with different screen sizes. In both experiments participants' performance was assessed with eye tracking recordings and visual and textual questionnaires.

As we have revised/discussed/commented on the results of both experiments in the previous sections, in these general conclusions we will consider the contributions of the whole study.

Even though the two experiments presented seem to focus on different areas, some core understandings can be drawn from them. From an attentional point of view a general pattern has emerged in the results of the two experiments carried out. If we go back to Kahneman's attentional theory, we can see the effects of the policy distributor and the evaluation of the demand mechanism working together closely in the processing of the information in the experimental situations. A modulation, or even a modification of the strategy we use to process information has been found in the two different experimental situations that we performed. According to the results, we can see how the effect that the attention modulation has is shown in the perception strategies used.

This change did not presented itself in the way we hypothesized. We had expected to find a difference in the total time spent on the two different areas, however, the difference was instead found in the way that this time is spent. As we have shown in the different experiments, the percentage of time spent in the subtitles area is about 35% versus 65% spent in the scene area. This percentage of time remained steady throughout all our experimental situations.

What we did find were differences in how this time within the area is spent. In both situations, we found differences in the number of fixations and/or the duration of the fixations. In line with the results of the study carried out by Bisson, Van Heuven, Conklin and Tunney (2014), the number of fixations and the duration of those fixations were the differentiating factor between the different experimental conditions. This seems to indicate a change in the processing strategies when we face situations where our resources are perceived to be insufficient.

Combining these findings with the answers to the questionnaires has given us another layer of information to explore. In experiment 2, we can see how the adapting mechanism is able to compensate for the different situations faced. The aforementioned change in the information recollection pattern seems to be enough to mitigate the smaller size of the stimuli. However, in experiment 1 we do see a loss of performance in the responses. The experimental manipulation concerning the different audio languages seems to trigger an adaptation of the motor response in an attempt to gather more information. This translates into differences in eye movements depending on the audio of the clip, we found a higher number of characters per fixation when the language is unknown. Nevertheless, this change in behaviour is not enough to compensate for the loss of an information channel (the audio) resulting in weaker performances, as can be see in the results of the questionnaires.

Recurring patterns emerge from both experimental situations. We can see that the gaze pattern changes in relation to the perceived effort of the task. In the experiment 1, we can see how the suppression of the known language (hence the suppression of an entire channel of information) triggers a change of eye movement behaviour. In experiment 2 we are triggering the same responses by modifying the size of the source of this information in a different way. This common procedure, seems to be a top-down response that is triggered to meet the exterior demands.

More work needs to be done to drive parts of this experiment to a more controlled situation. Subtitles present a very interesting case study from which to extract knowledge on how the attention modulates our search for information and how this same information is then stored. We are yet to find a way of standardizing the approach we take in these situations in terms of the stimuli we should present and the metrics we have to extract as well as how to extract them.

From our point of view, nowadays all the following experiments in this area must have some kind of scene control. As we have tried to make clear throughout this dissertation, in this situation the scene information is equally as important as the text. Future studies that evaluate only the textual information to assess the efficiency of subtitles are going to be seriously undermining the cognitive processes that underlie the processing of the information they are presenting.

At the same time, we can no longer consider non-relative eye tracking information as valid. In this experimental situation fixation count has to be linked to letter count. If we consider absolute values without compensating for the content that is being presented, we lose information as well as the capability to compare our results with other experiments. This is one of the main flaws that we can already acknowledge in our device comparison experiment. The novelty of performing studies with eye tracking on mobile devices involves some extra difficulties which have been resolved in the time since the data collection was completed.

Further research

From a psychological point of view, subtitles are a very interesting case study to understand the processes behind attention distribution. Plenty of manipulations to the subtitles can be done to study the effects they have on our abilities and resource distribution mechanisms.

For instance, one of the key factors in subtitling is a lack of correction during translation. The effect of changes in lexic can be very interesting on different levels. On an attentional scale, we might be able to see how the lack of coherence between the audio and the subtitles affects how we distribute our resources in the next subtitle presentation.

Another level of this experimental situation would be the effect this has on the understanding of the content. Does the word chosen by the translator improve our comprehension or make it worse? And in this last situation, would substituting the word for the one that was supposed to be there, according to the participants, eliminate the problem? More than just the usual translation debate underlies all these questions. Answering these questions would lead to a deeper understanding of how higher processes (detecting the wrong word when comparing it to the audio) affect resource allocation in a bottom down process.

Another area that needs to be explored is orientation response. Do different levels of comprehension of the audio channel affect the way we direct our attention in the very early stages of the perception process? If so, is it gradual or does it depends on proficiency?

Finally, a standardized methodology is needed in order to make sense of all the studies conducted using this methodology. As we have already pointed out, transforming eye tracking data by relating it to the content that is being displayed is an option. More control over the way performance is measured is also necessary. In order to gain this control both the screen and subtitle area must be taken into consideration.

From our point of view, more studies using subtitles from a cognitive psychology background should be carried out as this gives a fascinating insight into some of our perception and attentional mechanisms which are yet to be fully understood. This thesis has addressed very few of the possibilities it has put forward but with these last words we want to leave as many open questions as possible for those who are to come after us.

Bibliography

Asociación para la investigación de medios de comunicación (AIMC) (2016). *Enquesta General de Medios (Octubre 2015 – Mayo 2016).* Madrid: Asociación para la investigación de medios de comunicación.

Bisson, M. J., Van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). The role of verbal and pictorial information in multimodal incidental acquisition of foreign language vocabulary. The Quarterly Journal of Experimental Psychology, 68, 1306–1326. doi: 10.1080/17470218.2014.979211

Broadbent, D. E. (1958). Perception and communication. London: Pergamon Press.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct Measurement of Cognitive Load in Multimedia Learning. Educational Psychologist, 38(1), 53–61. http://doi.org/10.1207/S15326985EP3801_7

Cook, C. I. (2014). Netflix: A stepping stone in the evolution of television (Master Dissertation). University of South Florida St. Petersburg. Retrieved from http://dspace.nelson.usf.edu/xmlui/handle/10806/11681

D'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. European Psychologist, 12(3), 196–205. http://doi.org/10.1027/1016-9040.12.3.196

d'Ydewalle, G. Van Rensbergen, J. & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In O'Reagan J.K & Lévy A. (Ed.) *Eye Movements: From*

Physiology to Cognition (p. 313-321). Amsterdam: Elsevier Science Publishers B.V. (North-Holland).

De la Fuente, & J. Pousada, A. (2005). L'Atenció. Barcelona: UOCedita.

Deutsch, J. A. & Deutsch, D. (1963). Attention, some theoretical considerations. Psychological Review, 70, 80-90.

Diao, Y., Chandler, P. & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. American Journal of Psychology, 120(2), 237-261.

Díaz, J. & Fernandez, M. (2008). Using subtitled video materials foreign language instruction. In Díaz, J. (Ed.) The Didactics of Audiovisual Translation (p. 201-214). Amsterdam/Philadelphia: John Benjamins.

Díaz, J. & Remael, A. (2007). Audiovisual Translation: Subtitling. Manchester: St. Jerome.

Duchowski, A. T. (2002). A breath-first survey of eye tracking applications. Behavior Methods, Research, Instruments and Computers, 1, 1-15.

Duffy, S. A. (1992). Eye Movements and Complex Comprehension Processes. In K. Rayner (Ed.), Eye Movements and Visual Cognition: Scene Perception and Reading (p. 462–471). New York, NY: Springer New York. http://doi.org/10.1007/978-1-4612-2852-3_28

Gottlieb, H. (1998). Subtitling. In Baker, M. (Ed.) Routledge Encyclopedia of Translation Studies (p. 244-248). London/New York: Routledge.

Hayati, A. & Mohmedi, F. (2011). The effect of film with and without subtitles on listening comprehension of EFL learners. British Journal of Educational Technology, 42(1), 181-192.

HbbTv.org. (n.d.). Retrieved December 22, 2016, from https://www.hbbtv.org/overview/#how-it-works

Henderson, J. M., & Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In Underwood, G. (Ed.), Eye Guidance in Reading and Scene Perception (p. 269-293). Oxford: Elsevier.

International Television Expert Group (2013). One television year in the world. Retrieved from: http://international-television.org

Ivarsson, J., & Carroll, M. (1998). Subtitling. Smrishamn, Sweden: TransEdit HB.

Just, M. Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. Psychological Review, Vol 87(4), 329-354. http://dx.doi.org/10.1037/0033-295X.87.4.329

Kahneman, D. (1973). Attention and Effort. Englewood Cliffs, NJ: Prentice-Hall, INC.

Kastrenakes, J. (2015). The iPad's 5th anniversary: a timeline of Apple's category-defining tablet. The Verge. Retrieved from: http://www.theverge.com/2015/4/3/8339599/apple-ipad-five-years-old-timeline-photos-videos

Kruger, J.-M. Esté, H. & Matthew, G. (2014). Attention distribution and cognitive load in a subtitled academic lecture: L1 vs. L2. Journal of Eye Movement Research, 7(5), 1–15.

Lavaur, J.-M. & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to the viewers' fluency level and to the language in the subtitles? International Journal of Psychology, 46, 455–462. doi:10.1080/00207594.2011.565343

Loftus, R. (1972). Eye Fixations and Recognition Memory for Pictures. Stanford University Academic Press. 525–551.

Loftus, G. R. (1981). Tachistoscopic simulations of eye fixations on pictures. Journal of Experimental Psychology Human Learning and Memory, 7(5), 369–376. http://doi.org/10.1037/0278-7393.7.5.369

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. Journal of Educational Psychology. 90(2), 312–320. doi:10.1037//0022-0663.90.2.312

McLeod, P. D. (1978). Does probé RT measure central processing demand? The Quarterly Journal of Experimental Psychology, 30, 83-89.

Media Consulting Group. (2007). Study on Dubbing and Subtitling Needs and Practices in the European Audiovisual Industry. Final report. European Commission, Directorate- General Education and Culture.

Morrison, R. E., & Rayner, K. (1981). Saccade size in reading depends upon character spaces and not visual angle. Perception & Psychophysics, 30(4), 395–396. http://doi.org/10.3758/BF03206156

Morton, J. (1969). Interaction of information in word recognition. Psychological Review, 76(2), 165–178.

Neves, J. (2008). 10 fallacies about subtitling for the d/Deaf and the hard of hearing. The Journal of Specialised Translation, 10, 128-143.

Norman, D. A. & Bobrow, D. G. (1975). On data-limited and resource-limited process. Cognitive Psychology, 7, 44-64.

Comenius J. A. (1896). Orbis Sensualium Oictus.

Olsen, A. (2012). The Tobii I-VT Fixation Filter Algorithm description. Sweden: Tobii Technology.

Orrego-Carmona, D. (2015). The Reception of (Non) Professional Subtitling (Doctoral Thesis). Universitat Rovira i Virgili, Barcelona.

Paas, F. G. W. C., & Van Merrienboer, J. J. G. (1993). The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. Journal of Human Factors and Ergonomics Society, 35(4), 737–743. http://doi.org/10.1177/001872089303500412

Paas, F. G. W. C., Van Merrienboer, J. J. G., & Adam, J. J. (1994). Measurement of Cognitive Load in Instructional Research. Perceptual and Motor Skills, 79(1), 419–430. http://doi.org/10.2466/pms.1994.79.1.419

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P.W.M. (2003). Cognitive load measurement as a means to advance cognitive load theory. Educational Psychologist, 38(1), 63-71. doi: 10.1207/S15326985EP3801_8

Paivio, A. (1986). Mental representations: A dual coding approach. Oxford, England: Oxford University Press.

Paivio, A. (2006, October). Dual Coding Theory and Education. Paper presented at Pathways to Literacy Achievement for High Poverty Children. Michigan.

Perego, E., Laskowska, M., Matamala, A., Remael, A., Robert, I. S., Szarkowska, A. & Bottiroli, S. (2016). Is subtitling equally effective everywhere? A first cross-national study on the reception of interlingually subtitled messages. Across Languages and Cultures, 17(2), 205–229. http://doi.org/10.1556/084.2016.17.2.4

Perego, E. Del Missier, F. Porta, M. & Mosconi, M. (2010). The Cognitive Effectiveness of Subtitle Processing. Media Psychology, 13(3), 243–272.

Polder, E., & Parker, D. (1985). Evolution of Television Receivers from Analog to Digital. Proceedings of the IEE, 73(4), 599-612.

Posner, M. L. & Boies, S. J. (1971) Components of attention. Psychological Review, 78, 391-408.

Rayner, K. (1998). Eye movements in Reading and Information Processing: 20 Years of Research. Psychological Bulletin, 124(3), 372–422. http://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Smith, T. J., Malcolm, G. L., & Henderson, J. M. (2009). Eye movements and visual encoding during scene perception. Psychological Science, 20(1), 6–10. http://doi.org/10.1111/j.1467-9280.2008.02243.x

Robson, G. D. (2004). The closed-captioning handbook. Amsterdam, Netherlands: Elsevier.

Sadoski, M. & Paivio, A. (2004). A dual coding theoretical model of reading. In Ruddell R. B. & Unrau N. J. (Ed.), Theoretical models and processes of reading (5th ed.) (p. 1329-1362). Newark, DE: International Reading Association.

Sanders, A. E. (1993). Processing information in the functional visual field. In d'Ydewalle, G. & Van Rensbergen, J. (Ed.), Perception and cognition: Advances in eye movement research (p. 3-22). Amsterdam: North Holland. Sanders.

Subtitle. (n.d). In Oxford English Dictionary. Retrieved from: https://en.oxforddictionaries.com

Talaván, N. & Rodríguez-Arancón, P. (2014). The use of reverse subtitling as an online collaborative language learning tool. The Interpreter and Translator Trainer, 8(1), 84-101.

Talaván, N. (2010). Subtitling as a task and subtitles as support: pedagogical applications. In Díaz, J., Matamala, A. & Neves, J. (Ed.) New Insights into

Audiovisual Translation and Media Accessibility (285-299). Amsterdam/New York: Rodopi.

Tobii Technology AB. (2013). Tobii Technology AB Accuracy and precision Test report. Stockholm.

Tomaszkiewicz, T. (2006). Przekład audiowizualny. Warszawa: Wydawnictwo PWN.

Treisman, A. M. (1964). Verbal cues, language and meaning in selective attention. American Journal of Psychology, 77, 206-219.

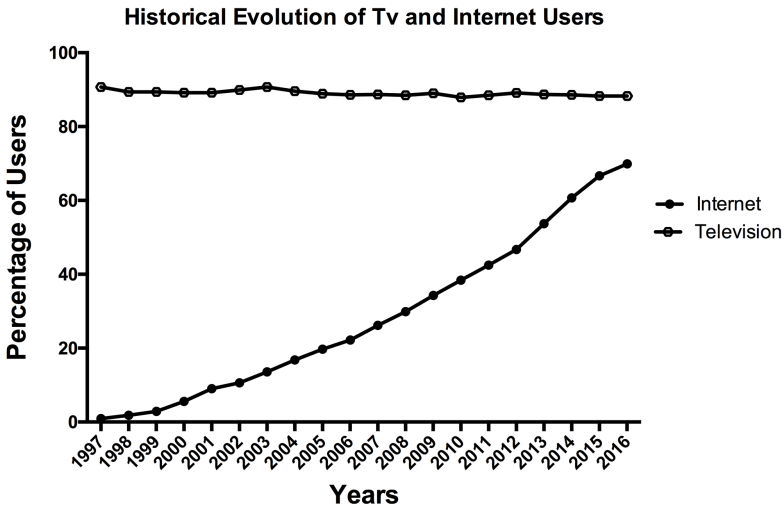
Vanderplank, R. (1988). The value of teletext subtitles in language learning. ELT Journal, 42(4), 272-281.

Vincent, J. (2017). The iPhone turns 10: a visual history of Apple's most important product. The Verge. Retrieved from: http://www.theverge.com/2017/1/9/14211558/iphone-10-year-anniversary-in-pictures

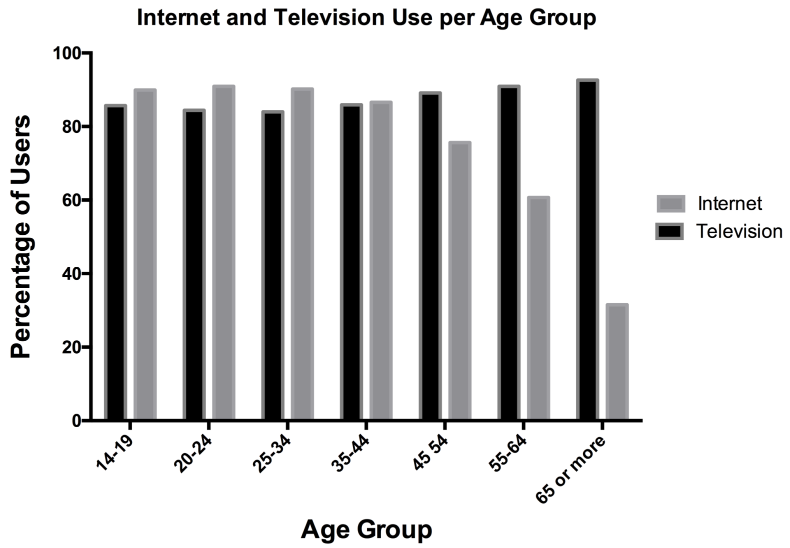
Zenith. (2016). Media Consumption Forecast (2016). London: Zenith.

Graphs & Figures

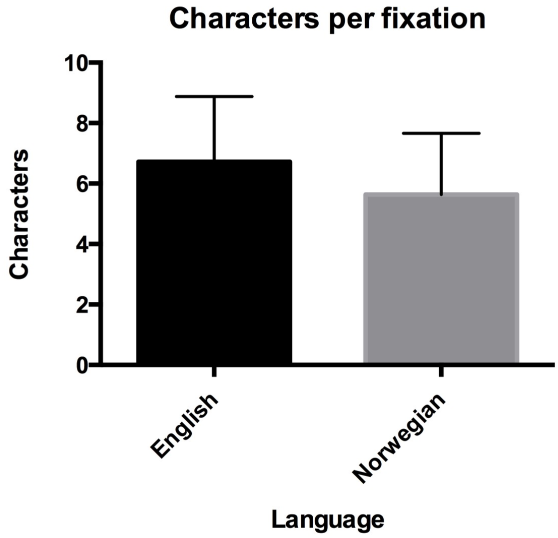
Graph 1: AIMC (2016). Historical penetration evolution 1997 – 2016 (%)



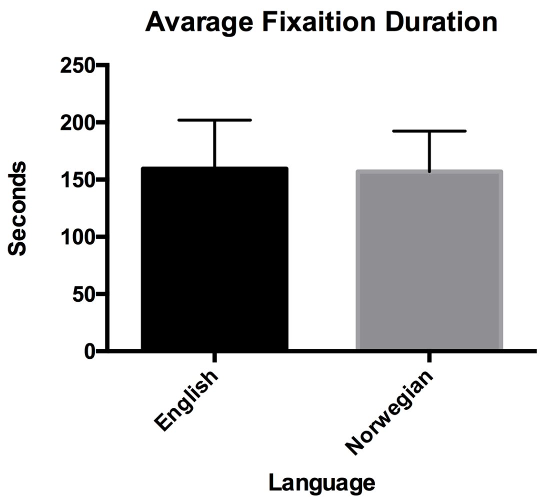
Graph 2: AIMC (2016). Internet use per age (data from May 2016).



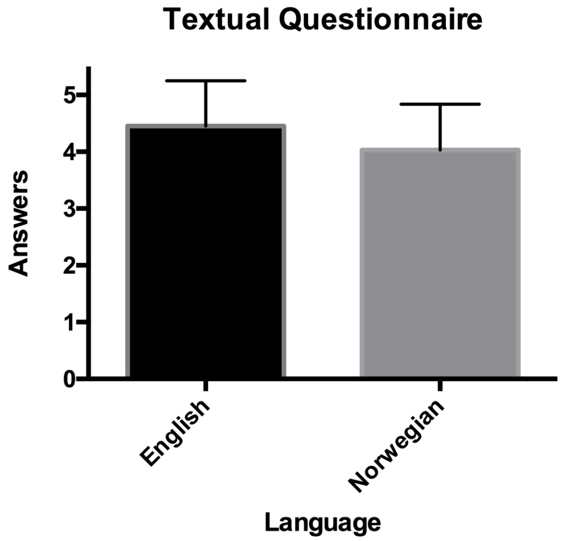
Graph 3: Experiment 1. Average Characters per Fixation.



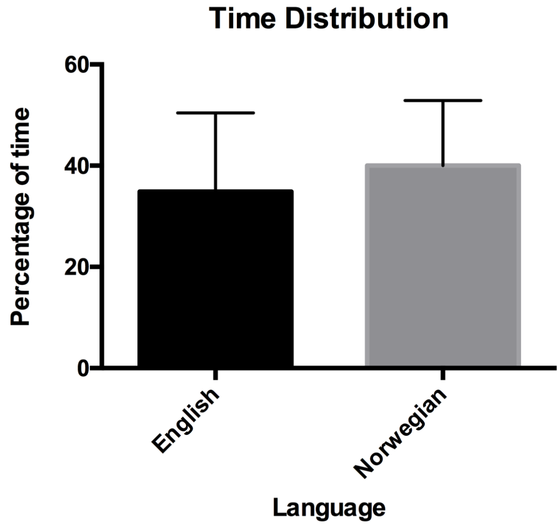
Graph 4: Experiment 1. Average Fixation Duration.



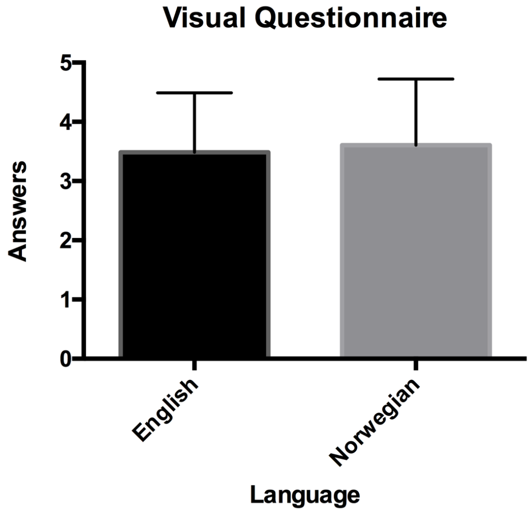
Graph 5: Experiment 1. Textual Questionnaire results.



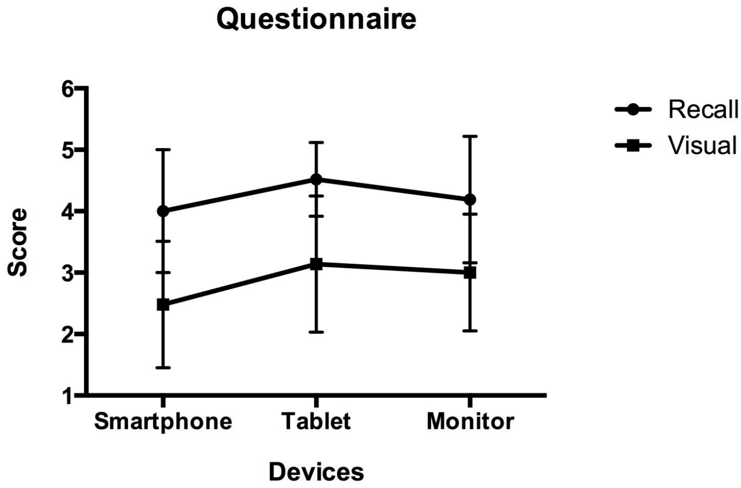
Graph 6: Experiment 1. Time Distribution Between Subtitles and Scene.



Graph 7: Experiment 1. Visual Questionnaire Scores.



Graph 8: Experiment 2. Visual and Textual Questionnaire Results by Device.



Graph 9: Experiment 2: Visit Duration (seconds) between areas on the different devices.

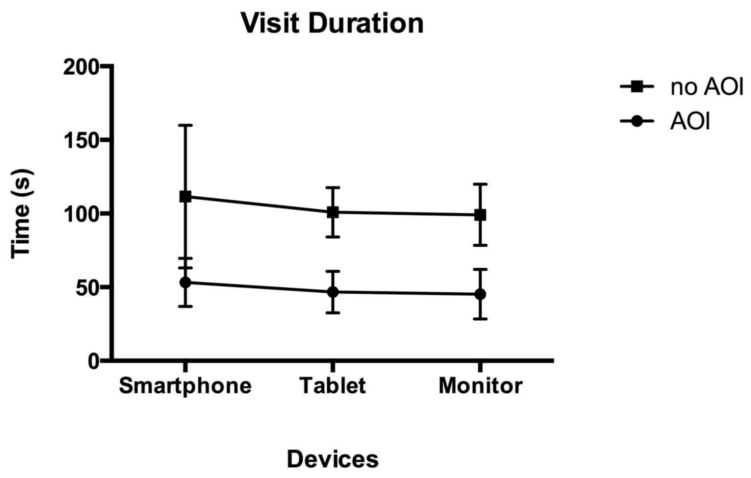


Figure 1: Paivio, A. (2006). Subsystem representation of the Dual Coding Theory.

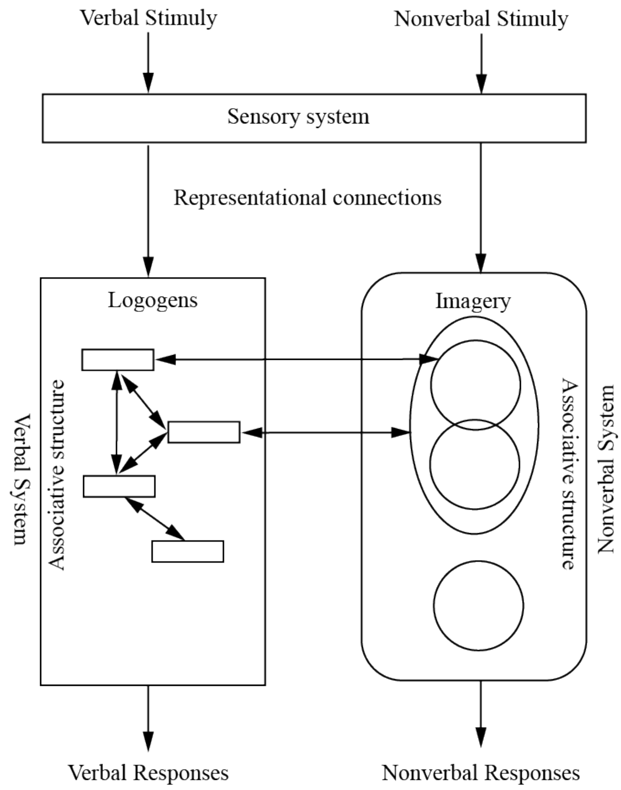


Figure 2: Mayer, R. E., & Moreno, R. (1998). A Dual Processing Model of Multimedia Learning.

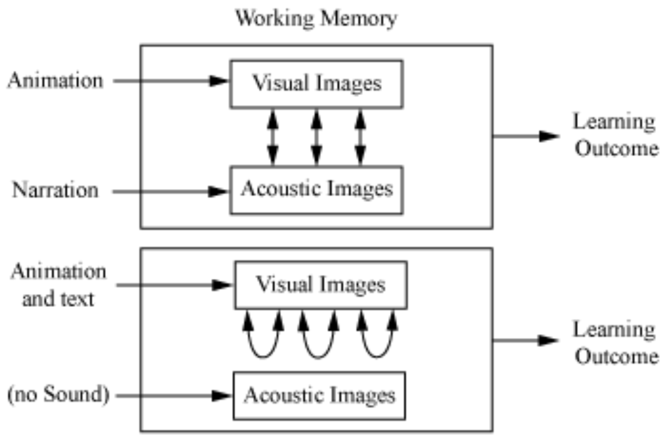


Figure 3: Broadbent (1958). Broadbent's ON/OFF Filter Model.

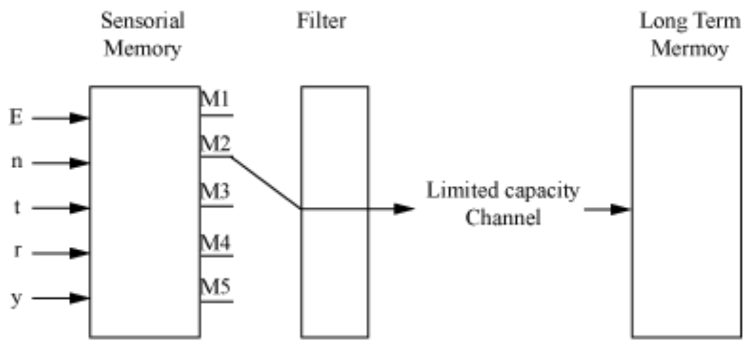


Figure 4: Deutsch and Deutsch (1963). Representation of the Post Categorical Filter.

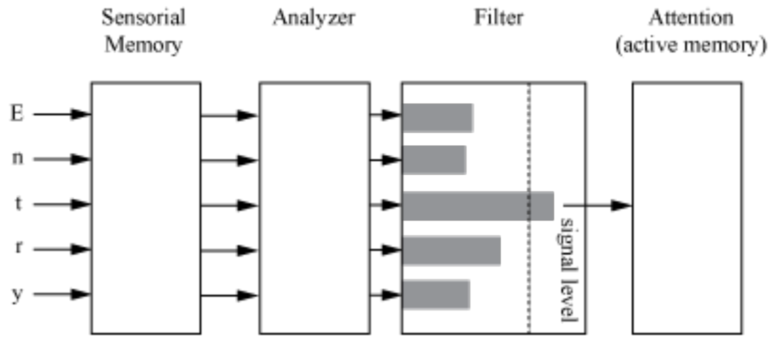


Figure 5: Kahneman, D. (1973). A Capacity Model for Attention.

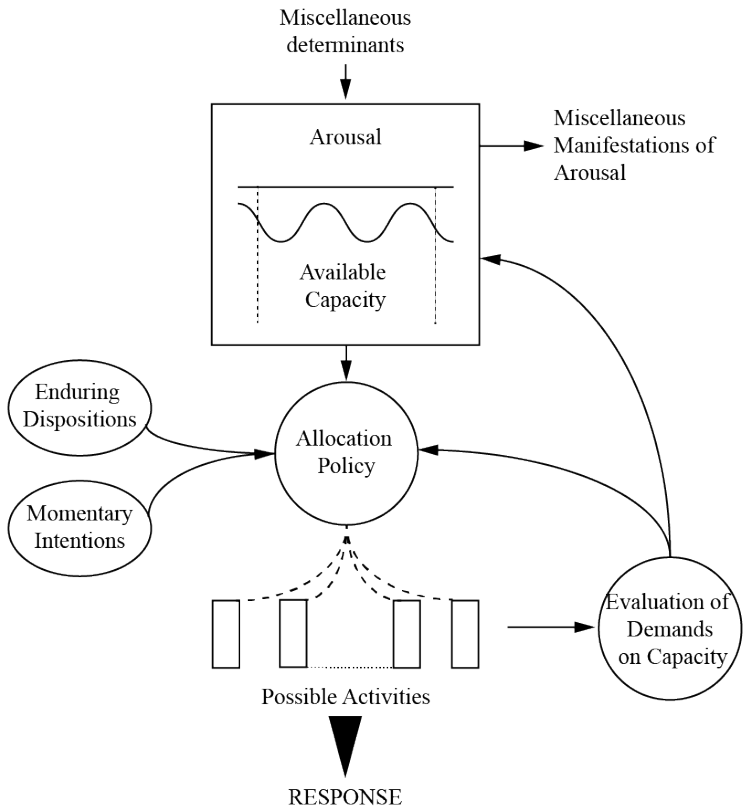


Figure 6: DCT and Kahneman attentional theory Integration Model.

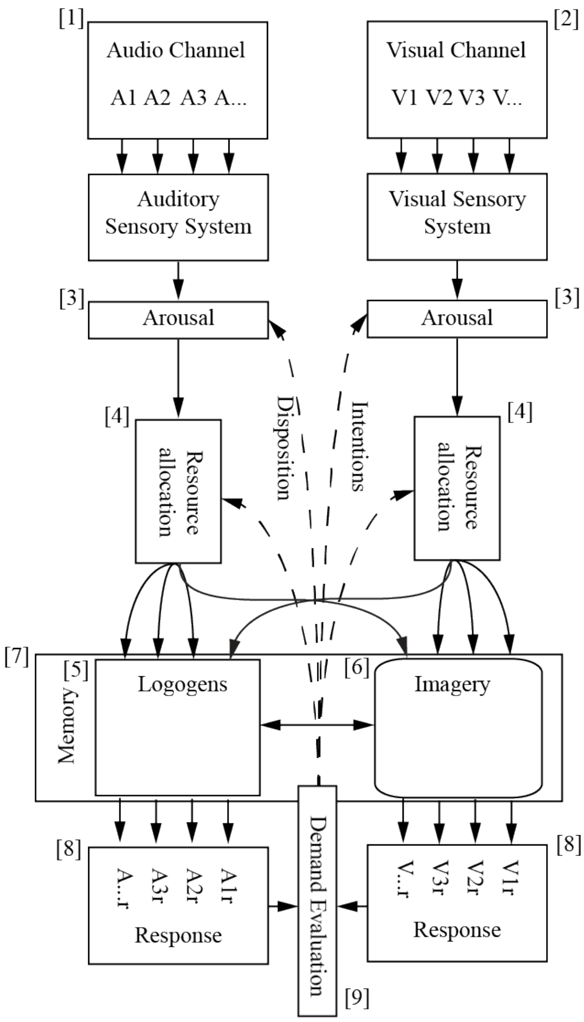


Figure 7: Application of the DCT and Kahneman attentional theory Integration Model to our experimental situation.

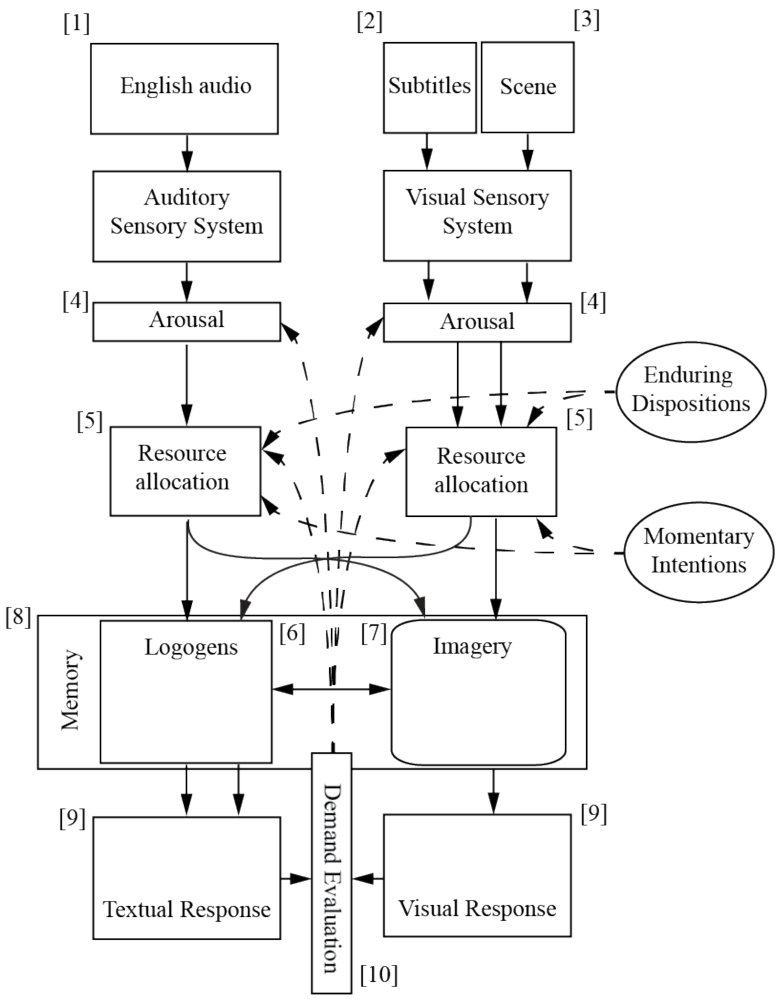


Table 1: Experiment 2. Mean Fixation Count Between Devices.

	Mobile	Tablet	Monitor
Subtitles	189.07*	203.27	201.6
Scene	260.6	281.73	310.33

* indicates a trend

Table 2: Mean Fixation Duration (in seconds) between devices.

	Mobile	Tablet	Monitor
Subtitles	.2387*	.1760	.1547
Scene	.3287	.2620	.2287

* indicates statistically significant differences

Table 3: Experiment 2. Time Distribution Between the Scene Area and the Subtitle Area on all Devices. In percentage of total time.

	Mobile	Tablet	Monitor
Subtitles	32.33	31.65	31.34
Scene	67.66	68.35	68.66