國立臺灣師範大學英語學系

碩　士　論　文

Master's Thesis

Department of English
National Taiwan Normal University

人人都需要"字幕"嗎?
探討不同學習行為
如何影響字幕對英語聽力理解之效度

Captions for All?
Validating the Effect of Captions on L2 Learners with Different Online Processing Profiles

指導教授：劉宇挺 博士

Advisor: Dr. Liu, Yeu-Ting

研 究 生：關蕙芯

Graduate: Kam, Emily Fen

中華民國一百零七 年 八 月

摘要

在多媒體學習中，字幕影片常用來訓練第二語言英語聽力。過去研究也發現字幕能協助第二語言學習者理解英語影片內容。然而，近年來有一整合分析發現，學習者不同的學習行為，會讓他們在理解字幕影片的英語內容時，受到字幕影響的程度也有所差異。此研究探討學習者的兩種學習行為—學習型態（聽覺與視覺型)和記憶能力(高或低)—如何影響字幕對英語聽力理解之效度。84名英語學習者的結果顯示，上述兩種學習行為對字幕的效度有不同程度的影響，甚至有顯著的交互作用。此研究發現學習者的記憶能力高度影響學習型態對英語聽力理解之效果。對記憶能力較差的學習者而言，他們的學習型態對英語聽力理解並無顯著的影響；然而，對記憶能力較佳的學習者而言，觀看無字幕的影片對聽覺型的學習者而言比較合適，反之，視覺型的學習者比較適合觀看有字幕的影片來增進聽力理解。本研究結果再次呼應不同學習行為對英語聽力的重要影響，特別是學習者如何透過字幕理解英語多媒體的資訊內容。同時，本論文也提出對教師、學習者與多媒體研發者對於英語聽力之建議，尤其針對不同學習行為提出適切的字幕影片以提升學習者的英語聽力。

關鍵字：第二語言英語聽力、多媒體學習、字幕、學習型態、記憶能力

Abstract

Captioned video is widely used by L2 learners to enhance their exposure to native oral input beyond the classroom setting. Such exposure to captioning has been found to provide useful visual aid for parsing and understanding L2 oral discourse. Notwithstanding, a recent meta-analysis has shown that captioning exerts a selective effect on L2 learners with different input processing profiles. This study investigated whether L2 learners' modality preferences (visual vs. auditory) and working memory capacity (high vs. low) would modulate the effect of captions on L2 listening outcomes. Results from 84 participants revealed that both cognitive variables impacted their L2 listening to different extents. Notably, working memory capacity modulates the impact of L2 learners' modality preferences on their listening outcome. Modality preference did not exert any significant impact on the listening outcome from those with lower working memory capacity. For the L2 learners with high working memory capacity, their modality preference was crucial in determining their listening outcome; in this case, visual learners did best when watching videos with captions, whereas auditory learners exhibited the best listening outcome with captions were not provided. The findings in this study further shed light on the importance of taking individual differences into consideration when employing captioned videos to maximize L2 learners' listening for different pedagogical purposes.

Keywords: L2 listening, multimedia learning, captions, modality preference, working memory capacity

Acknowledgment

This long journey of many discoveries would not have been completed without the kindness, support, and love from many.

To Tony:

We both know this finished piece did not happen without your unconditional and loving support. Words cannot express the emotions I have towards everything you have put into this journey. Thank you Boss.

To my blood-related family:

Daddy, Mommy, Michael, I love you, thank you for everything, seriously.

To my friends at NTNU:

Winnie, Jojo, Eric, Aletha, Wendy, Tiffany, and many more. Life at NTNU would have sucked so bad if you guys were absent. Thank you all for this exhilarating ride. Special thanks to Winnie. Thank you for bringing me home.

To my soul family—our home:

The moment I stepped into our home my soul had never cheered this loud for 3000 something years. You are the backbone, the fire, the waves, the breeze, the ocean, the forest, the light, and the love in my life.

Finally, I would like to dedicate this entire journey to Habala, Jesus, and the 12 Archangels. Thank you for your creation, love, and everlasting light. The road is still long, but faith will never fade.

At last, thank you, Master JC, for everything.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

Understanding running speech requires listeners to simultaneously process various input (e.g., audio text, facial expressions, and gestures) (Vigliocco, Perniss, & Vinson, 2014). Yet whether such processing can be successfully executed depends on how listeners hold (i.e., the functioning of working memory capacity) and select (i.e., online modality preferences—visual vs. auditory) relevant information delivered multimodally (different representations of the content presented). While this dynamic processing occurs automatically in first language (L1) listeners, it does not apply to all second language (L2) learners, especially those who have yet to automatize their oral decoding skills. And since human cognitive processing capacity is limited (Mayer, 2001), it makes automatizing L2 listening skills even more challenging.

To help overcome the aforementioned challenge, captions in videos have captured pedagogical interests as a multimedia scaffold to enhance the understanding of L2 oral discourse. Designed to help segment and parse running speech during listening, captions are visuals texts delivered simultaneously with animation that matches the auditory target language (Leveridge & Yang, 2013; Stewart & Pertusa, 2004). The addition of captions to videos allows L2 learners to simultaneously parse and process the discourse through dual or multiple channels, which is believed to enhance the width and depth of multimedia learning (Low & Sweller 2005; Mayer, 2001, 2005a; Moreno

2006; Moreno & Mayer, 2007). In fact, the benefits of captioned videos have been evidenced in some empirical studies, showing that captioning provides instantaneous, useful visual aid for parsing. Specifically, captioning is found facilitative to beginning and intermediate L2 learners' understanding of L2 running speech (Chai & Erlam, 2008; Gruba, 2004).

Although some studies have endorsed the pedagogical potency of captioning on L2 listening, the findings from a recent meta-analysis study (Perez, Noortgate, & Desmet, 2013) illuminated that captioning may exert a selective effect on L2 learners with different proficiency profiles (i.e., beginning, intermediate, and advanced). That is, participants' different degrees of L2 command could *modulate* the effect of captioning on their listening comprehension. Nevertheless, existing studies that have attempted to further this line of research have yielded inconclusive results due to methodological differences (e.g., achievement tests, proficiency tests, length of study, and other individually certified criteria) in determining participants' L2 proficiency levels (Leveridge & Yang, 2013; Pujola, 2002; Taylor, 2015). Such methodological incongruence poses comparability issues in existing L2 caption research, raising questions as to whether the findings can be generalized to L2 learners of different proficiency profiles.

The above inconclusive findings of L2 caption research on listening comprehension have jointly implied that not all L2 learners would benefit from captions equally. The issue at focus narrows down to: who will? To further explore this issue, it is imperative to target variables that *directly* affect real-time listening process. Thus far, L2 caption research has yet to explore factors that influence how L2 listeners simultaneously process information presented through multiple channels. Whether these online input processing profiles alter the effect of captioning awaits empirical investigation, which could potentially shed light on optimizing its effect on L2 listening comprehension.

In light of the significance of online input processing profiles, this study selects two variables that are highly relevant to the multichannel processing of L2 listening: L2 learners' working memory capacity (Mayer, 2005a) and their preferred input processing modality (Leveridge & Yang, 2013). With multiple input (i.e., video, audio, texts) at play during video watching, how efficient L2 learners process these information at their disposal depends on their working memory capacity (Mayer, 2001; Moreno & Mayer, 2007). And since L2 learners are limited in the amount of information they can process, it is conceivable that those with better working memory capacity may benefit more from the addition of an extra input—caption—while viewing video. Furthermore, the constraint of processing capacity forces L2 listeners to be more selective in their attention allocation (Mayer, 2005b), which is likely to drive L2 learners to orient their attention to the information delivered in their most preferred modality (e.g., visual vs. audio). Nonetheless, whether both variables indeed modulate the effect of captioning in a multimodal environment is yet to be empirically established. Further investigation is needed to offer evidence-based guidelines when using captioned videos to enhance listening comprehension for L2 learners of different input processing profiles.

To the above end, the current study is set out to explore whether differential working memory capacity and modality preferences during input processing would modulate the effect of captioning on L2 listening comprehension. By investigating how learners with different L2 online processing profiles benefit from captioned videos, the present study provides suggestions regarding differentiated instruction on the use of captions that caters to individual differences during multimodal input processing. In language learning classrooms, in particular, providing students with different avenues of learning allows them to comprehend audiovisual materials according to their preferred mode of learning. This way, learning becomes personalized even in a classroom setting, which in turn, empower all students to learn effectively.

As captioning is found not beneficial to all L2 learners, it is crucial for instructors to gain further insights on learner differences to provide optimal multimedia scaffold for L2 listening. The two research questions explored in this study are as follows:

1. Does working memory capacity modulate the effect of captions in understanding L2 videos? If so, to what extent?
2. Does preferred input processing modality (i.e., visual and auditory) modulate the effect of captions in understanding L2 videos? If so, to what extent?

CHAPTER 2

LITERATURE REVIEW

This chapter reviews theoretical frameworks and empirical findings that are crucial to identify the research niche and contextualize the present study. To this end, the literature review will consist of five major sections. The first section (2.1) aims to provide the theoretical background for this study, and (the theoretical framework) will introduce Mayer's cognitive theory of multimedia learning (CTML)—a theoretical framework addressing how human minds process multiple input in a multimedia-enhanced learning setting. This framework is then discussed vis-à-vis the use of captions to enrich learners' L2 listening experience. In particular, three key assumptions of CTML that address the success and constraint of multimodal input processing will be highlighted, followed by a discussion vis-à-vis the effect of captioning on L2 listening.

After addressing the theoretical basis of using captioning to assist L2 listening, the second section (2.2) will review existing L2 caption research. Throughout a critical review of these empirical studies, the third section (2.3) proposes that captioning exerts a selective effect on learners with different L2 proficiency profiles, which further validates the necessity of investigating factors highly relevant to real-time parsing. In this vein, the fourth (2.4) and fifth (2.5) sections discuss two input processing factors that might modulate the effect of captioning in assisting multimodal (L2) listening process: 1) working memory capacity; and 2) preferred modality in real-time input processing (visual and auditory), respectively.

2.1 Theoretical Underpinnings of Using Captions to Assist Multimodal L2 Processing

L2 Listening entails rigorous real-time processing of input (i.e., audio input of speech, prosody, postures, gestures, facial expressions, and body movements) delivered from different channels (Vigliocco et al., 2014). Such real-time processing requires efficient and flexible attentional control, allowing L2 learners to attend to various input in order to parse running speech and form mental notes (Vandergrift, 2004; Vanderplank, 2010). This attentional allocation, however, is not always easy for L2 learners—especially those whose L2 is still in progress—for they are limited in the amount of information they can rehearse and process in the working memory (Mayer, 2001, 2005a). As the above cognitive constraint increases the difficulty of automatizing oral decoding and encoding skills, it makes multichannel listening a cognitively challenging activity for many L2 learners.

To overcome the aforementioned challenge, instructors and researchers have sought various multimedia supports to facilitate L2 listening. Among them, captioning in videos prevails as an increasingly popular listening scaffold. Captioning—the visual texts delivered with corresponding oral discourse—has been extensively used to facilitate and enhance L2 learners' listening experience, for its presence gives additional cues for L2 learners to better visualize, parse, and hence understand the running speech. Although such additional input may impose cognitive load to L2 listening process, its support has been found and established in many empirical evidence—captioned videos could potentially aid L2 listening comprehension (e.g., Garza, 1991; Guichon & Mclornan, 2008; Hayati & Mohmedi, 2011; Huang & Eskey, 1999; Markham, Peter & McCarthy, 2001; Perez et al., 2013; Winke, Gass, & Sydorenko, 2013).

The multimodal effect of captioned videos is theoretically premised on a "learner centered" and "cognitive-constructivist oriented" theory—Mayer's (2001, 2005a) cognitive theory of multimedia learning (CTML). CTML stipulates that multimedia designed to invoke processes compatible to how the brain works is more likely to lead to meaningful learning. In this vein, any multimedia-enhanced learning that provides and calls for *multimodal* processing would potentially facilitate L2 *multimodal* process. This contention thus offers the theoretical basis for using captions as a way to enrich the multichannel input provided by video to enhance L2 *multimodal* listening.

Although CTML believes that multimedia-enhanced learning (MEL) has the potency to enhance L2 *multimodal* process, Mayer argues that this stipulation may not be true in all cases. He proposes three assumptions that would either facilitate or constrain learners' multimedia-enhanced processing experience—dual channels, limited capacity, and active processing assumptions.

First, regarding the dual-channel assumption, Mayer and other cognitive scientists argue that the outcome of multimodal processing is optimal when learners are simultaneously exposed to verbal (i.e., spoken or written words) *and* nonverbal input (such as animation, picture, or video), rather than just either input alone. Specifically, when both input are simultaneously presented to learners, they are more likely to build elaborate, meaningful and coherent mental representations of the input (Mayer, 2001, 2005a, 2005b). He further posits that such simultaneous processing of various verbal and nonverbal input encourages active mapping and integration with prior knowledge, which is likely to generate greater depth of learning (Low & Sweller, 2005; Mayer, 2001; Moreno & Mayer, 2007). The above dual channels assumption is theoretically based on the architecture of working memory. Specifically, working memory entails rehearsal/processing mechanisms for visual and auditory input, respectively. That is, what is heard (i.e., audio text and background sounds) will be processed in the auditory phonological loop and what is seen (i.e., animation and written texts) will be processed in the visuospatial sketchpad. Multimodal input that combines both verbal (e.g., captions) and nonverbal (e.g., video content) has the potency to activate the functioning of visuospatial sketchpad *and* phonological loop, thereby optimizing the processing in working memory.

Second, despite the benefit of simultaneous processing of multimodal (e.g., verbal + nonverbal) input, working memory has its limitation. This gives rise to the second assumption in CTML: the limited capacity assumption. In simultaneous input processing, working memory functions as a middleman that first converts sounds and images from sensory memory to verbal and pictorial models, and later integrates the converted models with prior knowledge before storing in long-term memory (see Figure 1). During the above processes, whether simultaneous processing of multimodal input would be beneficial to learners depends on whether the aforementioned middleman job can be efficiently and automatically performed by learners.

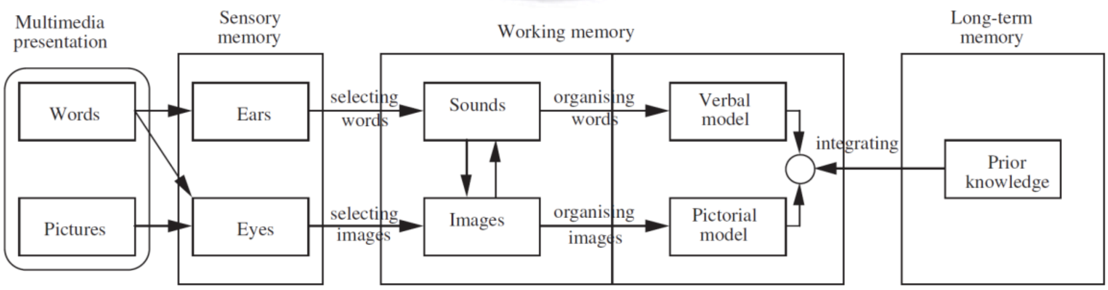

Figure 1. Mayer's cognitive theory of multimedia learning (Mayer, 2010, p. 545)

The benefits of the above processing, however, may not apply to all learners for working memory capacity varies significantly from person to person. Since such cognitive capacity is in charge of the amount of information learners can temporarily rehearse and retain in each channel at a given time, its variance may become a crucial determiner to their learning success. In this case, when a captioned animation or video—a typical multimodal input—is presented, L2 learners are only able to attend to limited portions of the video input (e.g., the video content without the captions) in their working memory rather than snap shooting the entire (video) content (captions + video content). Any cognitive overload from either channel (i.e., visual or auditory; verbal or nonverbal) can jeopardize learning for working memory capacity is exceeded (Jong, 2010).

In light of inter-learner difference in their processing capacity, Mayer (2001, 2005a) proposed the third assumption of CTML: the active processing assumption. Specifically, due to limited processing capacity, learners have to be highly active and selective in determining what is relevant and what to attend to among different modalities (i.e., videos, audio, and captions) in order to learn meaningfully and to benefit from MEL (Mayer, 1996, 1999). This active role learners play during MEL emphasizes the importance of learner control (Hasler, Kersten, & Sweller, 2007) in selecting input during real-time multimodal processing. Since learning is personal and invisible to the eye, individual differences in how learners utilize different input is likely to influence their MEL outcomes. It is conceivable that learners may rely more heavily on the input from the modality they prefer in multimodal processing. In the case of using captioned videos to assist L2 listening, which modality of input L2 learners prefer the most and select during multimodal listening is likely to modulate the extent to which they will attend to captions, and ultimately benefit from it. But when the aural input is preferred over the visual input, the advantage of captioning in aiding multimodal L2 listening could be compromised. Accordingly, whether or not L2 learners can take captioning support to their listening advantage depends on their working memory capacity and their preferred modality in multimodal input processing.

The theoretical underpinning from CTML helps the present study contextualize the potential role of captioning in enhancing L2 listening; importantly, the effect of captioning in assisting L2 multimodal listening may be modulated by differential learner input processing profiles, in particular with regard to their working memory capacity and preferred modality in input processing (see more elaboration on this issue in subsection 1.4-1.6). However, this has not been examined in existing L2 captions studies. To contextualize our examination of the role of working memory capacity and preferred modality in modulating the effect of captions, a review of existing captions studies is warranted. The following section will review L2 caption studies investigating the effects of full/partial captioning on L2 listening comprehension.

2.2 Empirical Studies of Captioning on L2 Listening

Defined as bimodal (audio and visual) presentation of texts delivered via videos in multimedia, captioning is believed to facilitate the development of various listening skills to enhance L2 listening. When input is enhanced through the provision of captioning, it makes indistinct auditory text salient to L2 learners (Chang, Tseng & Tseng, 2011; Danan, 2004). Such saliency helps visualize word boundaries and parse the aural streams, making multimodal input more comprehensible (Bird & Williams, 2002; Garza, 1991; Winke, Gass, & Sydorenko, 2010). It also provides appropriate scaffolding to L2 learners when input is slightly beyond their proficiency level (Danan, 2004), and thereby increases learner motivation by reducing anxiety during listening (Vanderplank, 2010). Accordingly, captioning facilitates the process of decoding and understanding L2 content, and has the potential to generate a positive mindset towards L2 listening (Bird & Williams, 2002).

With the prevailing belief that captioning benefits L2 listening, many empirical findings obtained from experimental designs have also lent support to such effect . An earlier quest made by Garza (1991) used full captions in videos to investigate whether advanced L2 learners of Russian and English would achieve higher listening comprehension. The results showed that caption group significantly outperformed their counterparts in the content-based comprehension test. Specifically, when captions were provided, L2 learners of Russian had higher gains on comprehension than L2 learners of English. Similar benefits from captioning are also shown in Huang and Eskey's (1999) study. This study exhibits that intermediate ESL leaners who have watched television series with full captions scored significantly higher in the listening comprehension test. In a more recent study, Winke et al. (2010) asked a group of 2nd and 4th year Spanish learners to watch three selected videos twice with or without full captions. Then L2 listening comprehension was measured by using multiple choice tests. A significant difference was found between caption and no caption group, suggesting that the exposure to captions substantially aided the understanding of audiovisual materials. The findings are also in line with Mayer's CTML, which stipulates that when multimodality reinforces one another during input processing, the impact it has on L2 listening is substantial.

Other lines of research that compare full and no captions with other caption variations also support the prominence of captioning. Several studies have found that L2 intermediate learners with the support of L1 subtitles and L2 captions significantly outperformed the no-caption group in listening comprehension tests (Markham et al., 2001; Hayati & Mohmedi, 2011; Markham & Peter, 2003). In a more recent study, Mirzaei, Meshgi, Akita, and Kawahara (2017) compared the effect of two captioning conditions i.e., full (FC) and partial synchronized (PSC) with no captioning on 58 Japanese learners of English with varying command of the target language. The quantitative results corroborate with the previous findings (Markham et al., 2001; Markham & Peter, 2003; Winke et al., 2010), suggesting that the presence of captions, regardless of either FC or PSC, significantly aids listening comprehension.

On a larger scale, Perez et al. (2013) conducted a meta-analysis with an aim to determine the overall effectiveness of full and partial captioning on L2 listening. Although a large effect of full captioning on L2 listening comprehension was found (Hedges' g = 0.99), it was noted that different L2 proficiency profiles could account for variation of the effect size (beginning: $g = 0.6$; intermediate: $g = 1.6$, $p < .001$; advanced: $g = 0.7$).[1] Differences in participants' proficiency profiles may lead to comparability issues, as L2 learners of various command in the target language process input differently (Leveridge, 2015). Consequently, this difference may potentially modulate the effect of captioning. Further insights into how differential L2 proficiency profiles may adjust the effect of captioning and other MEL devices are warranted. This insightful research attempt will help explain why the effect of captioning is not equivocal—an issue to be turned to in the following subsection.

2.3 The Role of L2 Proficiency Profiles on Captioning in L2 Listening

[1] Perez et al. (2013) noted that this finding should be interpreted with care due to unproportioned inclusion of studies (beginning: intermediate: advanced = 3: 6:1)

As shown in the meta-analysis by Perez et al. (2013), there is still a lack of consensus on the effect of captioning, for it may not be facilitative to L2 learners with different proficiency profiles. In an effort to bridge this gap, Taylor (2005) investigated whether differences in the length of L2 (Spanish) study would modulate the effectiveness of captioning on L2 listening comprehension. Two groups of Spanish beginners were divided into captioning and no captioning session, and the data was divided into first-year learners versus 3-year learners. *t* test results from between-group comparison revealed that L2 learners with longer (3-year) study outperformed those in their first year in the captioning session. Importantly, between-group comparison from first-year learners showed that not only did captioning not aid their listening comprehension, it was in fact detrimental to their understanding of the audiovisual materials. This significant difference (p = .03), however, was not shown in third-year learners (p = .49), indicating that captioning was neither harmful nor beneficial to their listening process. Retrospective data subsequently revealed that captioning was more distracting to first-year students when attending to all three modalities (i.e., video, audio, and captions), yet all learners expressed positive attitude towards visual texts. Findings from Taylor (2005), therefore, not only sheds light on the possible modulating effect of different L2 proficiency profiles, but also raised an issue as to whether, with longer L2 exposure, captioning in videos can be a valuable scaffold to comprehension for beginners.

Contrary to the findings in Taylor (2005), Lwo and Lin's (2012) study found that captioning is in fact more beneficial to the less competent L2 learners. This study examined whether different target language competency (determined by their achievement test results) would benefit from different types of captioning (i.e., Chinese captions, English captions, Chinese plus English captions, and no captions). Results indicated that when target language competency was considered, the less competent benefited from viewing only English or Chinese plus English captions, whereas their counterparts did not receive any facilitative effect in the listening comprehension results. Such findings also correspond with the study by Chang et al. (2011), as it established that employing captions improved low English proficiency learners' listening comprehension as compared to their more advance counterparts.

Mixed results from the above studies echoed Markham's (1989) findings, suggesting that captioning support may benefit L2 learners of different proficiency profiles to different extents. However, findings on this difference have been anything but consistent; exactly which group of learners (e.g., advanced vs. beginning) would benefit the most from captioning is still an unresolved issue (Winke et al., 2010). Which learner profile factor actually modulates the effect of captioning requires more empirical validation for two important reasons. First, there is a lack of methodological homogeneity in determining the level of L2 command (e.g., achievement tests, proficiency tests, length of study, and other individually certified criteria) in existing studies. Such inconsistency does not shed light on which demographic factor(s) actually modulate(s) the pedagogical potency of captioning. Importantly, the inconsistency in methodological instruments poses comparability issue as to whether the results are comparable and generalizable to L2 learners of different proficiency profiles. Second, idiosyncratic L2 command obtained through *offline* measurement (e.g., achievement tests) – as seen in the studies reviewed above in this subsection – may not reflect their online information processing *while* viewing captioned video, which significantly underlies the nature of L2 listening. In this respect, factors that are more straightforwardly related to this process are more likely to determine the degree to which learners utilize captions in viewing videos. L2 learners' input processing profiles, therefore, may provide invaluable answer to unravel this quest.

The present study, therefore, selects two variables that are highly relevant to the nature of L2 listening—the functioning of working memory capacity and idiosyncratic modality preferences—to investigate the extent to which both variables modulate the effect of captioning. First, since listening is transient in nature, how automatic L2 learners simultaneously decode and encode multiple input will determine the amount of information that can be stored in their working memory for comprehension purposes.

Second, the multimodality of captioned videos requires efficient input selection; in this case, L2 learners' preferred input modality will most likely affect their caption viewing behaviors, and ultimately influence its listening effect. Therefore, to optimize the effectiveness of captions, more research investigating the aforementioned two factors is warranted.

2.4 Working Memory and its Potential Role in Modulating the Effect of Captions

Working memory, a cognitive system that is limited in maintaining and storing information in the short term, is a multicomponent concept that underlies human thought process in learning (Baddeley, 2003). Such information processing is not only important in L2 learning, but also essential to comprehending and producing language (Miyake & Friedman, 1998). In particular, this cognitive capacity has been established as a major source of individual differences in L2 learner profiles (Révész, 2012). As a result, the role of working memory capacity has received considerable attention in bettering the understanding of L2 learning.

Among the many theoretical models of working memory, Baddeley and Hitch's (1974) proposed model is the most widely accepted and cited. Its cognitive architecture consists of an *executive control* which subsumes two slave systems: (1) the *phonological loop* and the (2) *visuo-spatial sketchpad*. While the phonological loop is responsible for storing and processing verbal and acoustic messages, the visuo-spatial sketchpad is in charge of information concerning visual and spatial elements. The executive control operationalizes complex cognitive activities such as focusing and allocating attention; activating, inhibiting, and selecting different processes; and policing information traffic between short-term storage and long term memory.

The commonality that weaves the three components together is that they are all limited in capacity. How efficiently each system can complete a task largely depends on whether they can deal with other activities that need to be complete. For example, during listening, since articulation takes place in real time, it increases the items that need to be rehearsed in phonological loop. In this case, the more incoming items, the more likely the first item will decay before it can be rehearsed (Révész, 2012). The second principle of Mayer's CTML—Whether or not listeners are able to process multi-sensorial and multimodal input efficiently and with automaticity—is based on this theoretical premise regarding the constraint of the phonological loop and the visuo-spatial sketchpad.

In the multimodal context of captioned videos, such cognitive capacity may become a determinant to whether captioning is facilitative to the depth of listening comprehension. Because such activity takes place under time pressure, it is conceivable that greater working memory capacity enables more efficient processing of multiple input. Empirical evidence has shown that individuals with higher scores in working memory test—an indication of having greater working memory capacity—have better control over what they attend to, exhibiting more dexterous skills during information processing (Colflesh & Conway, 2007). More control in processing indicates more flexibility in utilizing verbal information, and thereby enabling deeper input processing to occur (Révész, 2012). Consequently, L2 learners are better able to retain the comprehended materials for longer period of time. Hence the effect of captioned videos is highly dependent on the cognitive capacity of working memory in maintaining and processing multiple incoming items. And since such limited capacity (Engle, Kane, & Tuholski, 1999; Mayer, 2001,2005a, 2005b) varies from one person to another, it can then be hypothesized that individual differences in working memory capacity is likely to lead to differential degree of listening success during real-time information processing. In light of the aforementioned rationale, more research on the mediating role of working memory capacity in L2 listening is needed in order to optimize the effect of captioning in videos.

In the following section, we will discuss another important factor that may modulate the effect of caption, preferred input modality (i.e., audio and visual) in online input processing. As noted above, whether multimodality in captioned videos is positive or negative to L2 listening may depend on L2 learners' input modality predilections during real-time input processing.

2.5 Preferred Input Modality and its Potential Role in Modulating the Effect of Captions

To benefit from captions in multichannel listening, L2 learners have to make the best of their attentional resources while actively process all three modalities (i.e., audio, video, and texts). The biggest challenge lies in the fact that, as pointed by Taylor (2005), L2 learners need to efficiently process input from different modalities, and then integrate it with existing schema or knowledge (Al-Shehri & Gitsaki, 2010). What L2 learners would draw on to a greater extent during this multimodal process is further complicated by L2 learners' idiosyncratic modality preference in real-time processing; L2 learners are more likely to focus more on the channel of information that is more preferred for parsing and for comprehension (Leveridge & Yang, 2012).

The possibility that L2 learners may process input in light of their preferred modality implies that L2 learners would learn best when their preferred input modality matches with what is presented to them (Oxford, 2003). Under this circumstance, visual learners may comprehend better under captioned viewing condition due to their reliance on the visual support of captioning. Auditory leaners, on the other hand, may not benefit as much from under captioned viewing condition because captioning (presented visually) may not be of their preference during input processing. In this respect, their real-time processing may not be affected by the presence or absence of captioning. Additionally, different modality preferences may lead to excessive information processing from the preferred channel, which could consequently, impose extra load on L2 learners' limited input processing capacity. Differential modality preferences, therefore, should be considered as an important factor that could modulate the benefits of captioning, which may in turn, exert various listening outcomes.

To optimize the effect of captioning, an accurate understanding of L2 learners' caption viewing behaviors is helpful, especially in the case of exploiting different input combinations to benefit L2 listening (Leveridge & Yang, 2012; Sun & Dong, 2004). Nevertheless, questions regarding how L2 learners of different modality preferences process captions or what they attend to have gone unanswered (Winke et al., 2010). To uncover L2 caption viewing behaviors during listening, an understanding on learners' visual and auditory preference for processing linguistic input is needed. Thus far, many researchers have employed retrospective techniques such as questionnaires or self-reports (see Reid, 1995 , for examples of such surveys; see also Scarcella & Oxford, 1992) to gain an understanding of L2 learners' self-perceived modality reliance. However, results from offline measurements may not accurately reflect on the online functioning of input processing, which is often left unopened to L2 learners' conscious state of mind during listening (Leveridge & Yang, 2013; Winke et al., 2013).

To accurately assess visual-auditory preferences in language learning, Leveridge and Yang's (2013a, 2013b) caption reliance test (CRT) offers a time-sensitive assessment that provides a relatively reflective and accurate measure of L2 learners' real-time modality preferences. In their study, the CRT was used to assess EFL high school learners' reliance on captions while viewing captioned videos. The results exhibited individual variance in the degree of reliance, with lower-level learners relying more on captions than their higher-level counterparts. Also, a negative correlation was found between caption reliance and their listening comprehension scores; EFL learners with a heavier reliance on captions (visual learners) tended to score lower in their listening comprehension tests. The insights from this study, therefore, demonstrates the intimate relationship between modality reliance and L2 listening, which again stresses its importance in affecting the effect of captioning in multichannel processing.

Furthermore, the application of CRT also helped Liu and Todd's (2014) study to identify preferred learning styles (i.e., visual and auditory) as a modulating factor in determining the effect of dual-modality input in enhancing L2 learners' reading comprehension and vocabulary acquisition. The study found a selective effect of dual-modality input on L2 learners' performances, with modality preferences as a key mediating variable. With both studies demonstrating the importance of differential preferred modality, the reliability of CRT is confirmed as a plausible measurement to activities that are carried out in real-time manner. In this vein, the present study administers the CRT test to determine L2 learners' preferred modality, and explore how this individual variance modulates the effect of captioning on L2 listening in a multimodal environment (i.e., captioned video).

2.6 Summary

This literature review has provided a critical appraisal of theoretical underpinnings and empirical evidences to explore how the effect of captioning may be potentially modulated by L2 learners' working memory capacity and modality preferences. To summarize this study, the listed followings are the main ideas based on the studies reviewed above:

1. Captioning (full or partial) in videos is a multimodal processing experience that has been theoretically motivated and empirically established to enhance L2 listening comprehension.
2. Nevertheless, existing L2 caption studies show that captioning exerts a selective effect on L2 learners with different proficiency profiles, and it is still unclear as to which learner profile factor actually modulates the effect of captioning.

3. Since the modulating factors investigated in the previous studies were not directly related to *how* L2 learners actually process input during listening, different input processing profiles, such as differential working memory capacity and input modality predilections, should be empirically investigated to optimize the effectiveness of captions.

Investigating how different input processing profiles may potentially affect the effect of captioning will shed light on optimal differentiated instruction when multimedia (in this case captioned videos) are used to assist L2 listening. This insight may help teachers tailor their instructions and/or use different video materials (with and without captioning) to meet different individual needs during their multimodal L2 listening. By understanding internal variance and its relationship with captioning support on L2 listening, the present study hopes to shed more light onto this issue in search of optimal listening outcomes in a multimodal-enhanced environment.

CHAPTER 3

METHODOLOGY

3.1 Participants

This study included 84 college-level learners of English from a local university in northern Taiwan. All participants were native speakers of Mandarin and had learned English as a foreign language (EFL) in their 12-year formal education. To ensure target language comparability, the participants were required to have a proficiency test (i.e., TOEFL, TOIEC, IELTS, or high-intermediate on GEPT) score that was equivalent to the B2 level of Common European Framework of Reference for Language (CEFR). Prior to the study, the participants were already familiarized with the functions and appearance of captioning support, for it is widely used in various English learning settings enhanced with multimedia.

3.2 Materials

3.2.1 Video Selection

The video used in this study was obtained from TED (Technology, Entertainment, and Design), a multifunctional platform that provides authentic listening materials. This feature is particularly useful for L2 learners who lack exposure to the target language outside of the classroom, such as the participants in this study. Hence, TED is considered a popular and suitable learning resource for L2 learners in an EFL setting. Another educational value of TED provides real-time captioning and oral transcripts. L2 learners are free to utilize these functions, making their multimedia learning experience more flexible and accessible. In a nutshell, the popularity, suitability, and accessibility of TED made it an ideal database for selecting video materials.

The aforementioned rationale propelled this study to select a TED talk video based on the following three criteria suggested in Mirzaei et al. (2017): (1) linguistic difficulty level (2) accent (3) topic familiarity. First, the linguistic difficulty of the video was compatible with the participants' actual proficiency level, for employing level-appropriate materials is crucial in generating reliable and valid L2 learning outcome in caption studies (Perez et al., 2013). Videos that contained low frequency words or academic specific terms were avoided to prevent misinterpretation or confusion (Mirzaei et al., 2017). Second, the speaker's accent in the video is standard American English, for it is the most popular and common accent used in the participants' English learning experiences. Third, the topic of the video had to be familiar to the participants, for L2 learners' background knowledge is found crucial in modulating the listening outcome (Bloomfield, Wayland, Rhoades, Blodgett, Linck, & Ross, 2010; Gebhard, 2000)

Adhering to the above three criteria, this study utilized a level-appropriate TED talk video on successful leadership—a topic closely familiar to the participants—as the viewing material. The speaker provided an analysis of the patterns collected from several successful leaders' effective communicative behaviors. Along with humor and character, the content of this video was considered captivating and educational for the participants in this study. The total length of the video is roughly 11 minutes, with average speech rate of 4.18 syllables/second. The video was played on a projector screen in a noise-free classroom setting.

3.2.2 Caption Viewing Conditions

Two caption modes used in this study were: full captions and no captioning. When the video was played, half of the participants received a full captioning of the corresponding oral text displayed at the bottom of the screen along with the video content, while the other half only viewed the video without captioning support. The video was played on an online viewing platform called *Youtube*, which provided the above two captioning options at viewer's disposal.

3.3 Design

A 2x3x2 factorial design was used in this study to explore the impact of two within-subject independent variables (working memory capacity and modality preferences) on one dependent variable (listening comprehension scores) under one between-subject independent variable (two caption modes). A randomized block design (see Figure 2) was employed to equally assign participants to the experimental (caption) and control (no caption) groups. By doing this, factor variability was controlled to ensure group

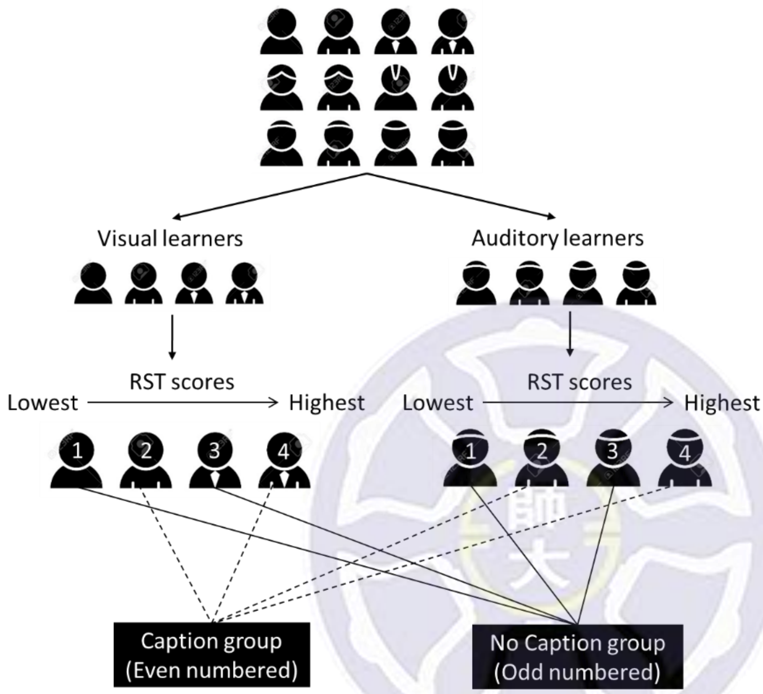

Figure 2. A visual schematization of the grouping procedure

comparability prior to the study. The grouping procedure began with all participants receiving (1) a CRT to measure their preferred input modality and (2) a reading-span task (RST) to determine their working memory capacity. After being assessed, all participants were first distributed into two groups based on their CRT results (visual or auditory learners). Then according to the ascending order (lowest to the highest) of RST scores, a number was assigned to each associated participant (e.g., the participant with the lowest RST score is number “1”, the second lowest is number “2”…). Once a number was assigned to each participant, the odd numbered were put into the no caption group, and the even numbered were put into the caption group.

As Figure 2 shows, this grouping process helped guarantee no statistical significant difference between the caption and the no caption groups in terms of their RST scores, which in turn, resulted in comparable data for later analysis. In addition, to make inferences about the modulating role working memory capacity played in this study, the participants were dichotomously distributed into two groups (i.e., high and low) according to the 50^{th} percentile converted from their RST scores. Scores below 50^{th} percentile were categorized as “low”, while scores above 50^{th} percentile were categorized as “high”.

Under both experimental and control conditions, the participants received a listening comprehension test to assess their understanding of the viewing materials. A questionnaire was distributed at the end of the study to collect information regarding the participants’ background information, their language learning experiences, and their perceptions on the tasks.

3.4 The Assessment Tasks and Scoring

The instruments that were employed in this study were: (a) the Caption Reliance Test (CRT); (b) a reading-span task (RST); (c) a listening comprehension test; (d) a post-study questionnaire.

3.4.1 The CRT

This study used the CRT to determine participants' modality preferences in real-time input processing based on two criteria: suitability and practicality. First, although the CRT was originally developed to measure L2 learners' reliance on captioning (Leveridge & Yang, 2014), it was later modified and used in Liu and Todd's (2014) study to determine L2 learners' modality preferences online (i.e., visual and auditory). Second, the items and length of the CRT can be easily adjusted to different kinds of test takers (e.g., an individual or groups of participants) as well as the testing environments (e.g., laboratory and classrooms). The CRT was, therefore, suitable for the investigation of this study and also highly practical for researchers and teachers with different user purposes.

The CRT consisted of items that involved two successive components: (1) watching a captioned video about a short dialogue between a man and a woman and (2) answering a multiple-choice comprehension question. The videos and questions were level-appropriate and suitable for the participants in this study. To prescribe modality preference online, the CRT featured congruent and importantly, incongruent items. In the congruent items, the audio texts and captions were identical; in the incongruent items, there was a one-word mismatch between what the participants heard and saw:

(The audio and captions from the dialogue below are simultaneously presented)

Men: Some of my students just won’t pay attention in class.

Woman: It might help if you introduce $\left\{ \begin{array}{c} \text{medication (what learners hear)} \\ \text{meditation (what learners see)} \end{array} \right\}$ to them.

As the example shows, participants simultaneously saw “meditation” and heard “medication” while viewing the captioned video. Under this circumstance, there was no right or wrong answer to the succeeding comprehension question, as both options appeared in the test item below:

Question: what can be the man’s solution to his problem?

(A) Meditation (B) Regulation (C) Medication (D) Recitation

Following Mayer’s CTML, it was assumed that visual learners were more likely to attend heavily to captions and thus gave answers based on what they see. In this respect, they were more likely to choose (A) from the example above. Auditory learners, on the other hand, may be more likely to select (C) for they preferred to rely more on input presented in sound form. To accurately assess preferred modality online, it was imperative that the participants were unaware of the incongruent items during the CRT.

To prevent this phenomenon from happening, this study adopted a 75% congruent and 25% incongruent ratio based on the guideline prescribed by Leveridge and Yang (2013). The participants' performance on the *incongruent items* was the focus of the succeeding analysis to determine their online modality preference.

Based on the scoring policy used in Liu and Todd (2014)—another study that also utilized the CRT—one point was awarded under the "visual modality" for choosing (A), and one point was awarded under the "auditory modality" for choosing (C). After totaling the points for either modality, each participant's preferred modality (PM) index value was calculated by dividing the visual modality points by the auditory modality points (see Table 1).

Table 1

CRT Score Ballet Sheet

	Visual modality Scores	**Non-target Scores**	**Auditory Modality Scores**	**PM index value**	**Preferred modality**
Participant # 1	7	0	5	1.4 >1	Visual
Participant # 2	3	0	9	0.3 <1	Auditory

The participants relying more on visual input, such as captions, achieved a PM index value larger than 1 (e.g., 7 visual modality points / 5 auditory modality points = PM index value > 1), while those attending more to the audio texts achieved a PM index value of less than 1 (e.g., 3 visual modality points / 9 auditory modality points = PM index value < 1). This categorization, consequently, divided participants into groups of two: (1) visual learners (PM value >1) and (2) auditory learners (PM value >1).

As for the outliers, if the participants chose answers from neither captions nor auditory text (answers other than "medication" or "meditation" from the above example), one point was given under the "non-target" answers. If there were two or more "non-target" points under the ten incongruent items, it suggested that the participants did not consistently choose the appropriate cues—either the auditory or visual (captions) cues—to comprehend the video content. This indicated that the CRT may be too difficult for the associated participants; in this case, their performance data were excluded from the analysis.

To ensure the content validity of the CRT, we have consulted three experts who are professional and experienced in the field of TESOL. Also, British National Corpus (BNC) was used to ensure the difficulty level of the CRT matched with the participants' L2. In addition, a pilot test revealed a close-to-even ratio between visual and auditory learners, which was almost identical to the ratio obtained from the actual results in this study. The identical ratio between the pilot and the experimental outcomes showed that the reliability of the CRT was secured.

3.4.2 Reading-Span Task (RST)

Many measures have been used to assess L2 learners' working memory capacity, but not all of them can provide a complete picture of such a 'dynamic' memory system. This system, according to Baddeley (2003), is responsible for storing, rehearsing, and simultaneously processing multiple input in real-time manner. To orchestrate the above dynamic processes, L2 learners have to rely on their executive (attentional) control *and* rehearsal mechanism, making both aspects of the functioning of working memory highly crucial in determining which measure to choose from.

To date, there are three working memory measures that were designed to tap into the functioning of the executive control and rehearsal mechanisms: counting span, operation span, and reading span tasks. In the current study, the reading span task (RST), originally developed by Daneman and Carpenter (1980), were used in consideration of practicality issues. Unlike the other two working memory assessment tasks, the RST can be easily implemented to a group of students in classroom settings (Turner & Engle, 1989), allowing instructors to efficiently make inferences about their students' working memory capacity for pedagogical purposes.

Typically administered on a computer, the RST had 12 trials in total. Each trial featured a sentence set consisted of 2-5 sentences. The RST used in this study included 3 trials consisted of 2 sentences, 3 trials consisted of 3 sentences, 3 trials consisted of 4 sentences, and 3 trials consisted of 5 sentences. Corpus of Contemporary American English (COCA) was used to determine the frequency of the sentence words to ensure that the difficulty level of the sentence words was appropriate for the participants in this study. All the sentences were composed of 13-16 words and were therefore, comparable in length (sentence length range: 12-17 words, 20-22 syllables, 55 to 73 letters). The maximum presentation time for each sentence was 6.5 seconds. All sentence sets were randomly presented to the participants on a computer screen to deconfound order effects (Engle, Cantor, & Carullo, 1992).

The first trial of RST began with the participants reading a sentence on the projector screen (e.g., *Emily slapped the face of a ridiculous sky*). Then a slide consisted of *yes* or *no* was presented to them to determine whether the sentence made sense or not (Turner & Engle, 1989). Immediately after responding *yes* or *no*, an isolated letter appeared in 1-second delay, while the participants attempted to retain the letter in their minds for later recall. After repeating the above procedure for all the sentences in a given trial, a question mark (See Figure 3) appeared on the screen to prompt the participants' recall of all the letters presented in a trial.

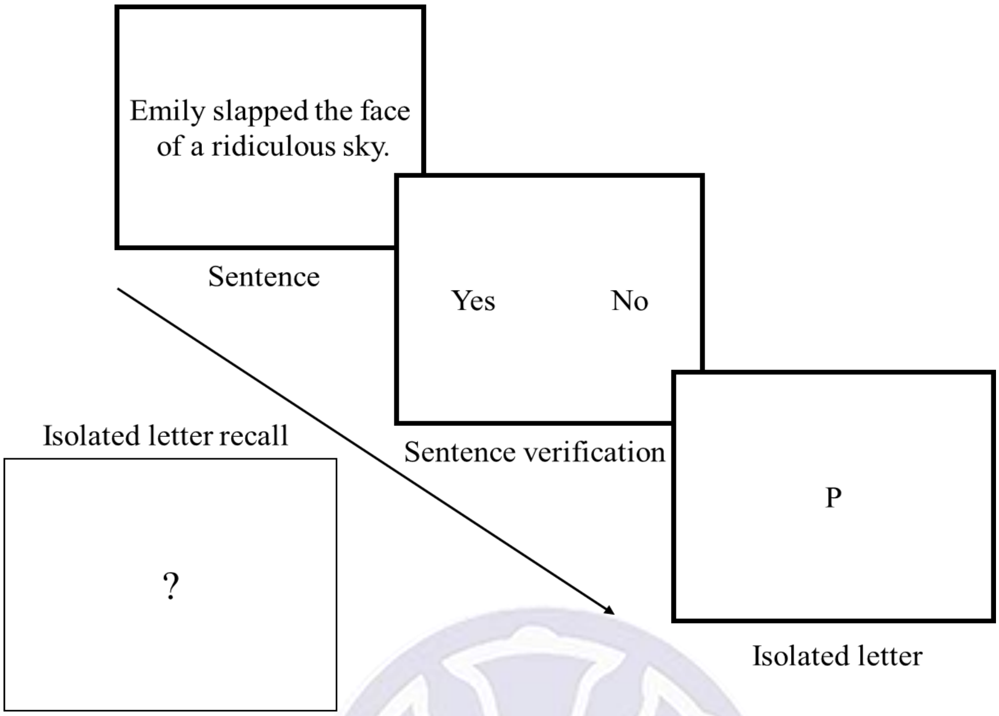

Figure 3. The procedure of a trial in the RST.

The first trial ended when the participants indicated that they had finished recalling all the letters before proceeding to the second trial. A paper-based instrument was used to collect all of the responses in this study. The entire task (12 trials) took roughly 20 minutes to complete.

To calculate the RST scores, the number of isolated letters recalled across all trials were totaled (one point for each accurate recalled letter) (Friedman & Miyake, 2004; Turner & Engle, 1989). The final scores allowed the researcher to make inferences about each participant's working memory capacity, the higher the better. Nevertheless, not all recalled letters were counted and included in the analysis. In order to ensure that the participants engaged in both storing (remembering isolated letters) and processing (understanding the sentences) – two critical functions of working memory – during the RST, this study adopted an 80% accuracy criterion from Turner and Engle (1989) for the sentence verification. In their study, those with total score below 80% accuracy rate were excluded from the analysis, for this suggested that the participants might be exclusively focusing on the isolated letters without devoting their attentional resources to processing sentences. This criterion, therefore, helped guarantee that the participants' attention was directed to the processing *and* the storing components in the task.

To ensure the content validity of the RST, professional suggestions from three experts in the field of TESOL were consulted. Also, British National Corpus (BNC) was used to ensure the difficulty level of the RST matched with the participants' L2. In addition, a pilot test was administered to establish the appropriate time lapse for sentence presentation and verification.

3.4.3 The Listening Comprehension Test

After watching the video, the participants—irrespective of their video viewing condition (with or without captioning support)—took a listening comprehension test designed based on the listening section of a TOFEL test, i.e., listening to the text provided and answering multiple-choice comprehension questions (see Appendix 2) on an answer sheet. This comprehension test contained 15 test items to assess participants' global, local, and inferential understanding of the content. Each item consisted of four options with only one correct answer.

The listening comprehension test was administrated in a classroom setting with the assistance of a computer, a projector, and a screen. The participants heard and read the stem of each comprehension question on the screen. They then read the options on the screen before writing their answers on the answer sheet. Each correct answer was given one point whereas an incorrect answer received no points. A pilot test took place to ensure the validity and reliability of the test items.

3.4.4 Questionnaire

The questionnaire (see Appendix 3) administered at the end of the study aimed to elicit participants' general experience of watching captioned videos and perception of watching the (captioned) video in this study. Clear instructions were given to ensure that all participants understood the meaning of the five-point Likert scale items.

Two versions of questionnaires were distributed to the no-captioned and full-captioned participants. For the no-captioned L2 learners, they were given the first version of questionnaire. In this version, the first section contained 8 items with an aim to collect the participants' general experience of watching captioned videos in their everyday lives. The second section contained 4 items to probe into the participants' video viewing experience in this study, with constructs such as, interest in the task, topic familiarity, level appropriateness of the task, and the degree of mental effort invested to complete the task. For the full-captioned L2 learners, the first section of the given questionnaire was identical to the no-caption counterparts. The second section, however, entailed 10 items to probe into their captioned video viewing experience. Specifically, the second section included items that attempt to elicit how the presence of captions affected the participants' processing of the multimodal input. At the end with an invitation to elicit participants' open-ended comments on their experience during task performance.

3.4.5 Interview

To collect participants' open-ended comments on their (captioned) experience during task performance, a follow-up interview was conducted after the participants filled out the questionnaires. Two professional experts were in charge of analyzing and interpreting the collected data with an aim to triangulate the questionnaire results.

3.5 Data Collection Procedure

Data collection began with gaining ethical approval, as all participants were asked to sign a consent form after reading the information sheet that explained the overall structure and procedure of the study, the potential benefits of participation, measures taken to guarantee their privacy and confidentiality, and the right to withdraw from the study at any given moment (see Appendix 1). Next, all participants were grouped into two viewing conditions based on their CRT and RST scores (see Figure 2). One week after the CRT and RST test, each group was separately invited to take part in a video viewing session. After viewing the video, a listening comprehension test was administered to assess their understanding of the viewing material. The session ended after the participants completed the post-study questionnaire followed by an interview session.

3.6 Statistical Analysis

STATISTICA 13.0 for Windows was used to generate descriptive statistics for the data in this study. A three-way ANOVA was used to explore the effects of the independent variable (i.e., caption modes) and the moderating variables (i.e., working memory capacity measure and modality preference) on the dependent variable (i.e., listening comprehension scores). The grouping conditions from this study resulted in sufficient number of participants per-condition to use three-way ANOVA to generate main effects and interactions among variables. Also, ANOVA was able to run data that contained one interval scale variable (i.e., listening comprehension) and three categorical scale variables (i.e., captions modes, working memory capacity, and modality preference). Independent *t*-tests were used to compare the means of different conditions. The alpha level of all tests was set at $p < .05$ as the level of significance for this study.

CHAPTER 4

RESULTS

This study was set out to investigate whether working memory capacity and online preferred modality affected the effect of captioning on L2 listening comprehension. The following sections will present the quantitative and qualitative data respectively. To begin with, an overview of the participants' captioning viewing behaviors will be disclosed throughout the descriptive statistics. Then the three-way ANOVA results will be displayed, along with the results from independent *t*-tests for post-hoc analysis. The questionnaire data supported by the interviews will be presented at last.

In this study, a total of 84 participants were recruited for the experiment. 12 of them were excluded from later analysis because they reported having already watched the selected (captioned) video. Additionally, they felt that content of the viewing materials was "too easy" for them. None of the above were desirable participant behaviors, which resulted in their absence in later data analysis.

4.1 Quantitative

4.1.1 Descriptive Statistics

Table 2 demonstrates an overview of the participants' listening comprehension scores vis-à-vis the factor/variable of this study. Although the mean difference between control (no caption) and experimental conditions (full caption) is small (M diff. = .005), larger variance is found in every condition with captions than without captions. Notably, such variance in the full caption conditions increases when taking individual differences (i.e., online preferred modality, working memory capacity, or both) into account. As such pattern consistently manifests throughout Table 2, it calls for a more fine-grained investigation on the analysis of the potential effects and interactions among the variables in this study.

Table 2

Descriptive Statistics of the participants' listening comprehension test scores.

Variables	Level of factors	N	Mean	SD
Total		60	11.15	1.62
CM	NC	28	11.17	1.46
	FC	32	11.12	1.77
PM	A	29	11.55	1.37
	V	31	10.77	1.76
WM	L	24	10.91	1.44
	H	36	11.30	1.73
CM*PM	NC*A	14	12.07	1.26
	NC*V	14	10.28	1.06
	FC*A	15	11.06	1.33
	FC*V	17	11.17	2.12

CM*PM*WM	NC*A*L	4	11.5	0.57
	NC*V*L	8	10.5	1.30
	FC*A*L	5	11.8	0.83
	FC*V*L	7	10.42	1.98
	NC*A*H	10	12.3	1.41
	NC*V*H	6	10	0.63
	FC*A*H	10	10.7	1.41
	FC*V*H	10	11.7	2.16

Note: Caption Mode (CM), Preferred Modality (PM), Working Memory (WM), No caption (NC), Full Caption (FC), Auditory (A), Visual (V), Low (L), High (H).

4.1.2 A three-way ANOVA

The Shapiro-Wilk test was used to determine the normality of the listening comprehension scores, which yielded a slightly left-skewed distribution (W=.958, p=.039). Since the results did not follow a normal distribution, this study adopted a non-parametric generalized technique (Thomas, Nelson, & Thomas, 1999) to convert the raw data into ranked data. This procedure allowed the research to perform normal parametric tests (e.g., t-test, ANOVA, and regression) on the converted comprehension scores.

The results from a three-way ANOVA test revealed that the inclusion of caption mode, online preferred modality, and working memory capacity can explain 27% variance of the listening comprehension performance (R^2=.27). Table 3 displays the main effects of the three independent variables on listening comprehension, showing that only the participants' preferred modality significantly affected their listening comprehension scores (p = .02). This significance was found to have a close-to-large effect size (η^2 = .09). Notably, the auditory learners significantly outperformed the visual learners in their listening comprehension outcomes, as Figure 4 visually schematizes such contrast.

Table 3

A three-way ANOVA result for the effects of captions modes, preferred modality, and working memory and their interactions on L2 listening comprehension

	DF	*F*	*p*	Partial eta-squared ($\eta^{2)}$	Observed power (alpha=0.05)
CM	1	0.44	0.5	0.008	0.10
PM	1	5.67	**0.02***	0.09	0.64
WM	1	0.001	0.96	0.00002	0.05
CM*PM	1	4.94	**0.03***	0.08	0.58
CM*WM	1	0	0.99	0	0.05
CM*PM*WM	1	5.797	**0.01***	0.10	0.65

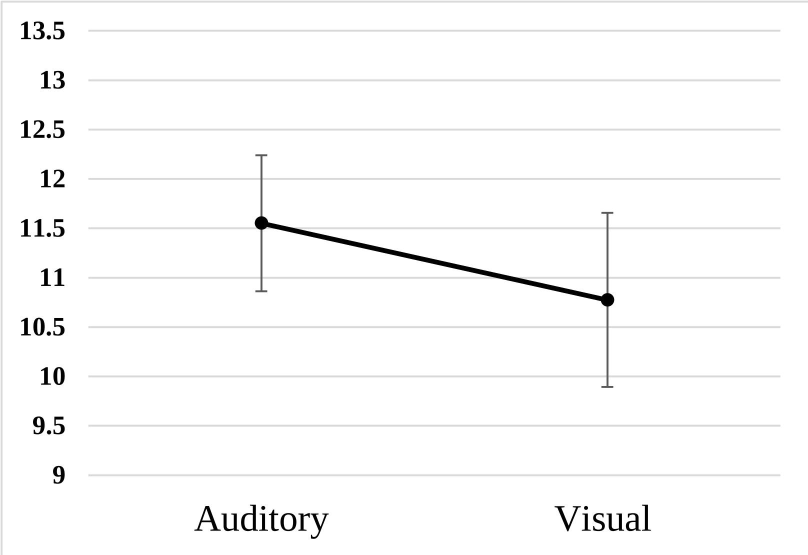

Figure 4. Mean difference between auditory and visual L2 learners' listening comprehension scores.

The listening advantage of the auditory learners, however, did not always hold true when taking caption mode into consideration. As shown in Table 3, there is an interaction (p = .03; η^2 = .08) between the participants' preferred modality and their caption viewing conditions (i.e., full caption or no caption) on their listening performances. Post-hoc analysis from an independent t-test further confirmed that captions had a significantly disruptive effect on the auditory learners (p = .048). The disruptive effect is visually schematized in Figure 5, showing that when exposed to captions, auditory learners performed substantively poorer than visual learners than without captions. However, without captions, the listening advantage for the auditory learners resurfaces, as they significantly outperformed their visual counterparts (*p* = .00) in the no caption condition. Despite the negative effect of caption imposed on the auditory learners, such an effect was not observed in the visual learners' listening comprehension under full and no caption conditions (*p* = .144).

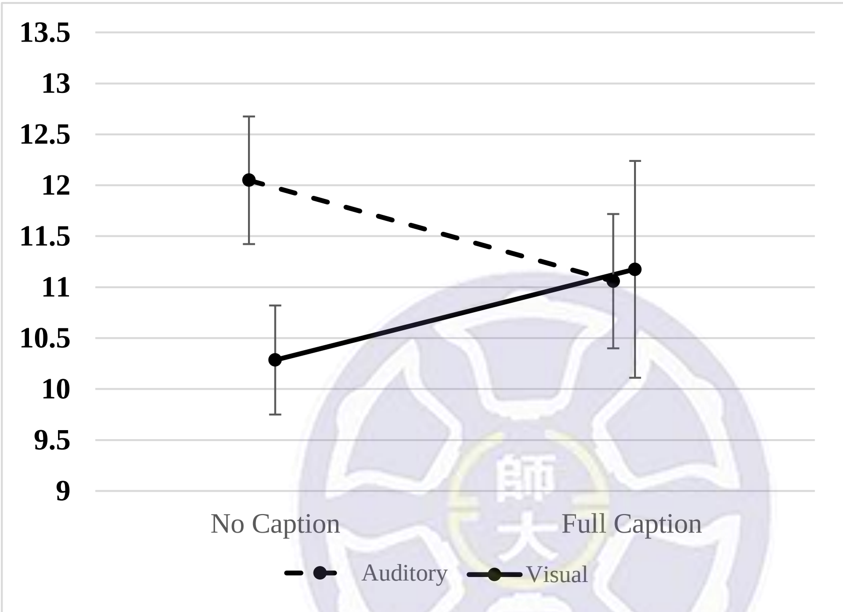

Figure 5. The interactions between caption mode and preferred modality.

Although the effect of captioning did not reach a significant level for the visual learners, a robust variance was observed in their listening performances (SD = 2.12). Such variance implies that preferred modality may not be the only factor that could potentially modulate the effect of captioning. This study found that whether captions were indeed facilitative to L2 listening comprehension was not only depended on preferred modality, but more importantly, working memory capacity.

As shown in Table 3, there is an interaction among caption mode, preferred modality, and working memory capacity (*p* = .01), with a close-to-large effect size (η^2 = .10). One may recall that the participants were grouped into having high and low working memory capacity based on their reading span test scores. Figure 6 shows that concerning the low working memory condition, neither did captioning exert a facilitative nor disruptive effect on both auditory and visual learners. That is, regardless of preferred modality, the presence of captions did not affect the listening comprehension of those with lower working memory capacity.

In contrast to the minimal effect captioning had on L2 learners with lower working memory capacity, a significant interaction was observed in the high working memory condition. Similar to the effects seen in Figure 5, the auditory learners with higher working memory outperformed the visual learners in the no caption condition (*p* = .001). When exposed to captions, however, it diminished the auditory learners' listening advantage, as they scored significantly lower than those in the no caption condition (*p* = .021). This result reinforces the detrimental effect captions had on auditory learners' understanding of the audiovisual materials, specifically those with higher working memory capacity.

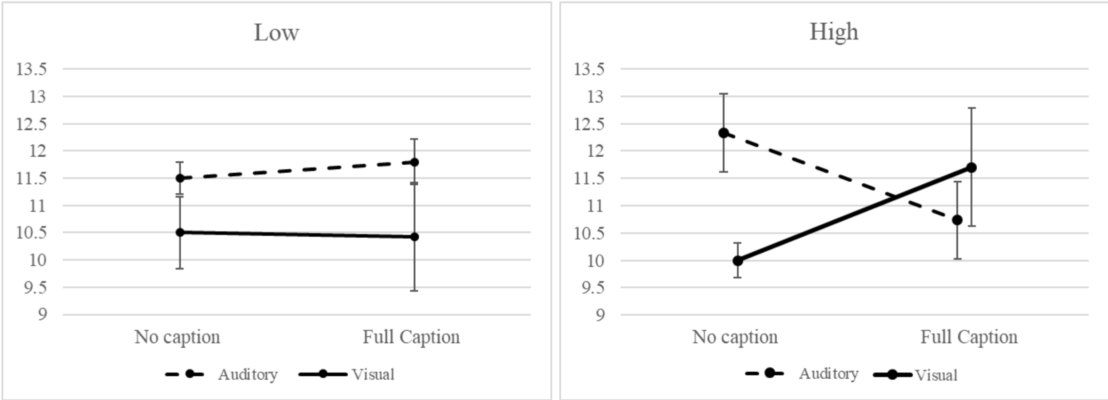

Figure 6. The interactions among caption mode, preferred modality, and working memory capacity.

Although captions failed to merit auditory learners with higher working memory capacity, it did wonder to their visual counterparts. Figure 6 illustrates that for the visual learners with higher working memory capacity, captioning significantly augmented their listening comprehension (p = .04)—a phenomenon not manifest in the visual learners' performance data when working memory capacity was not considered. When working memory capacity was reckoned, it accentuated the effect of captions. This phenomenon stipulates that the visual learners with higher working memory capacity may better process multimodal information with captions. The inclusion of working memory capacity, therefore, discovered the important role it has in determining the effect of preferred modality in affecting the listening outcome under different viewing condition (i.e., full and no captions).

4.2 Qualitative data

This study used both questionnaire and interviews as a secondary data source to support the quantitative results. Since the questionnaire was designed to probe the participants' perception/experience with captioned video content, the responses reported below will mainly focused on the participants assigned to the full caption (FC) condition. For the participants in the no caption (NC) condition, their perceptions toward captions will be reported by drawing on insights from their qualitative comments.

Table 4

Descriptive statistics of the questionnaire data with five-point Likert scale items (1 = Strongly disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly agree)

	Statements	**FC V**	**FC A**	**FC V H**	**FC A H**
		M			
Captioned video viewing behaviors	I focused more on the audio than the captions while watching the video.	2.41	4.6	2.7	4.93
	I focused more on the captions than the audio while watching the video.	4.47	3.57	4.87	3.87

Note. (1) FC = Full captions; NC = No captions; A = Auditory learners; V = Visual learners; H = High working memory capacity.

Auditory learners under the full caption condition

In Table 4, the questionnaire data showed that the auditory learners under the full caption condition reported that they tended to agree that their focal attention was directed to the oral discourse of the video (M = 4.6), in particular for those with higher working memory capacity (M = 4.93). Despite this, their average rating for the second questionnaire item ("I focused more on the captions than the audio") is 3.57, indicating that they tended to agree that despite their (auditory) modality preference, they were aware of their attention to captions. This is especially the case for those auditory learners with higher working memory capacity (M = 3.87). This finding is further supported by the qualitative comments from the auditory learners: two thirds of the auditory learners (8/15) from the full caption condition expressed that captions engaged a lot of their attention, four of which even showed strong dislikes towards captions. Additionally, when asked for their online-viewing preference, two thirds of the auditory learners under the full caption condition preferred not having captions on. Specifically, seven of them evinced that captions could direct their attention unwantedly, two of them found captions could be distracting, and 1 stressed that captions could impose more cognitive load to their understanding of the multimodal information. The followings are a few examples from the interview data:

"…Captions can attract a lot of attention from me, and I don't like it. When this happens, I feel like im not polishing my listening skills. I was paying a lot of attention on captions during video watching….”

“…Captions can be distracting…”

“…I fell captions could attract a lot of my attention during video watching…”

“…Captions can attract a lot of attention, and I don't really like my attention being controlled by captions…”

“…Even though I'd normally turn captions on, I still don't like my attention being controlled by captions…”

“…But with captions, it can impose more cognitive load because you have to constantly attend to so many things. It can also be a little distracting…”

<u>Visual learners under the full caption condition</u>

In contrast, the visual learners’ average rating for the first item is 2.41, indicating that they tended to strongly disagree that they attended to the oral discourse while viewing the captioned video; even for those visual learners with high working memory capacity (M = 2.7). When responding to the second item, the visual learners’ average rating is 4.47, suggesting that they tended to strongly agree that they attended to captions; this tendency is even more apparent for the visual learners with high working memory capacity (M = 4.87).

Unlike the auditory learners' negative comments noted above, visual learners exhibited a more positive attitude towards having captions during their online viewing experience. 6 of them explained that captions would not be distracting; 5 of them regarded captions as facilitative to their understanding of the audiovisual cues; 2 of them stressed that captions made them feel secure, allowing them to verify whether the received input is accurate. Specifically, having captions on would make them feel less cognitively demanded, for they intently relied on reading captions. The followings are a few examples from the interview data:

> "…I don't find captions distracting, it can help me understand difficult terms…"
>
> "…Watching video with captions is much easier (less cognitive load) for me; whereas without captions could be more demanding."
>
> "…Captions allow me to make sure that what I listen is correct. I really enjoy havnig captions because it allows me to verify whether my listening is correct…"
>
> "When I turn captions on, it's less cognitively challenging because I can just read them. Without captions, however, may be more cognitively challenging for me because I have to force myself to concentrate and to not space out."

Notably, five visual learners seem to treat captions as an optional asset, namely, an multimodal input that they could draw on when needed:

"…I feel I pay more attention to listening, and only briefly attend to captions when I need details to facilitate my comprehension…"

"Captions can be helpful for understanding the content. If I don't want to look at it, I just ignore it."

Auditory learners under the no caption condition

Seven auditory learners felt that captions were attention-demanding during video-viewing. They specifically concerned that captions directed their attention to specifics, rather than the tenor of the content. In that case, all of them preferred not having captions on during online viewing, for they concerned that captions were not conducive to their listening comprehension:

"…Captions may be helpful for understanding technical terms. I wouldn't turn captions on for this video because it's more focusing on understanding "central ideas"…"

"…I think if we were to train our listening comprehension skills, having captions on may be less desirable…"

"…Captions can draw a lot of attention and maybe this is not helpful for training L2 listening..."

Visual learners under the no caption condition

Twelve visual learners stated that they found captions attention-demanding, but not distracting. Importantly, five of them explicitly noted "I typically find captions helpful in promoting my comprehension, but viewing non-captioned videos (i.e., the TED talk video) also does not bother me at all because the video content already provides a lot of contextual visual clues." Notably, these comments all came from the visual learners with high working memory capacity.

CHAPTER 5

DISCUSSION

The two research questions addressed in this study intend to unravel whether L2 learners' preferred modality and working memory capacity modulate the effect of captioning. In particular, the extent to which the investigated variables interacted with each other will be discussed in the ensuing paragraphs, with an aim to further the understanding of captioning in the context of differentiated teaching/learning.

Prior to discussion, an interesting finding stood out. Specifically, this finding reinforced the need of considering L2 learners' input processing profiles (i.e., preferred modality and working memory capacity), especially when video instructional materials are used. Overall, the quantitative results showed that when L2 learners' input processing profiles were not taken into consideration, L2 learners—irrespective of their caption viewing conditions—did not differ significantly in their listening comprehension performances. The comparable outcomes from both caption and no-caption conditions indicated that the presence of captions did not significantly influence L2 learners' listening comprehension.

Overall, although the effect of captions was not significantly accentuated in this study, larger variance of listening scores was observed in the full caption condition ($SD = 1.77$). Notably, when L2 learners' input processing profiles were considered, such variance became greater, especially in conditions with full captions. The fact that larger variance was seen in full caption conditions indicated that the effect of captions may still exist, except selectively. Such selectiveness corroborated with a point made in chapter 2, suggesting that not all L2 learners would benefit from captions. Also pointed out in the previous sections, there is still a discrepancy regarding whether captions are truly beneficial to L2 listening comprehension. The ensuing sections, therefore, present the insights obtained from this study with an empirical quest to close the aforementioned discrepancy.

5.1 RQ1: Does preferred modality modulate the effect of captioning?

The first research question concerns whether participants' preferred modality (i.e., visual and auditory) would modulate the effect of captioning on L2 learners' listening comprehension. Statistical results revealed that this input processing factor significantly modulated the effect of captioning. It was also evidenced that modality preference alone resulted in statistically significant difference in the L2 learners' performance data. Both findings lent support to the hypothesis proposed in chapter 2, that modality predilections, as an input processing factor, could potentially modulate the extent to which the effect of captions is determined. Since preferred modality is theoretically and empirically pertinent to processing multimodal input, the following section will discuss the finding vis-à-vis both Mayer's CTML and existing empirical evidence from L2 caption research.

5.1.1 Effects of caption mode on auditory learners

The quantitative data revealed that L2 learners who were prescribed as auditory learners performed the best under the no caption condition. When compared to their captioned counterparts, the no-caption listening advantage became even more statistically salient. Therefore, it became clear that L2 auditory learners understood the audiovisual information best when immersed in no-captioned multimodal environment.

The prominent no-captioned effect manifested in the auditory learners may be accounted for by Mayer's active processing assumption in CTML. In theory, multimodal processing requires efficient learner control in selecting the most relevant input to aid comprehension (Hasler, Kersten, & Sweller, 2007; Mayer, 2001). Similarly, L2 learning context also requires learners to the same—be efficient in allocating attention to cues that are directly beneficial to listening comprehension (Taylor, 2005). In this case, L2 learners are most likely to depend on the input that matches their modality preference. Furthermore, Oxford (2003) hypothesized that in order to generate the most desirable learning outcome, what is preferred has to match with what is presented (Oxford, 2003). As the auditory learners in this study were found to be a case in point, it is reasonable to extrapolate that the no-caption condition may be a more favorable environment for them to optimize their multimodal listening outcome.

The observation that having no caption appears to be the optimal listening condition for the auditory learners indicates that visual support (in this case, any form of captions) disrupted, rather than facilitated auditory learners' understanding of multimodal video content. This explains why their comprehension was impaired when captions were provided. This finding contradicts with the previous studies, where captions were found to aid the listening comprehension process (Chai & Erlam, 2008; Danan, 2004; Markham & Peter, 2003; Sydorenko, 2010; Taylor, 2005; Vanderplank, 2010; Winke et al., 2010). In fact, prior research maintained that captions were widely used as a strategy to facilitate perceptual processing in listening (Goh, 2000). If captions were designed to facilitate the encoding of the acoustic message, why is it detrimental to the auditory learners' listening comprehension?

One possible explanation may lie in the disturbance of L2 learners' optimal processing channel. In the case of the auditory learners, their preferred processing channel is through what they hear rather than what they see. This speculation is established by the qualitative data reported earlier; namely when exposed to captioned video viewing environment, the presence of captions may have reoriented their attention to the visual input—the channel that is not of their processing preference.

Moreover, according to the questionnaire data, the auditory learners indicated that presence of captions may have divided the auditory learners' attentional resources. With limited attentional resources at hand (Mayer, 2001), it is possible that such demand from captions imposed difficulty in processing information from their optimal and most preferred channel. Furthermore, two thirds of the auditory learners did not express positive attitude towards receiving captions during their video-watching. Both sets of data sources, quantitative and qualitative, jointly indicated that the presence of captions may not lead to the optimal multimodal processing environment for the auditory learners. This finding, in turn, helped shed light on recalibrating the listening strategy suitable for those who preferred to "listen" in a multimodally enhanced environment.

5.1.2 Effects of caption modes on visual learners

Notwithstanding the notable effect captions had on the auditory learners, it did not manifest in the visual learners' listening outcome. This study found that the listening scores from the visual leaners in the caption condition did not differ significantly from those in the no caption condition. Without the assistance of captions, visual learners were able to achieve similar level of understanding as those assisted with the visualized text. This illuminating finding indicated that for visual learners, captions may not be the only scaffold facilitative to their listening comprehension.

Since visual learners tend to rely on what they see, it is possible that even without captions, they would still rely on other available elements presented visually. Nonlinguistic (e.g., facial expressions, body language etc.) input, for example, may contain subtle but rich contextual cues that may provide assistance equal to the effect of captions. As evidenced in Sueyoshi and Hardison's (2005) study, paralinguistic elements, such as gestures and facial cues, can assist the understanding of videotaped lectures—a type of video content similar to the TED video used in this study. This may explain that without captions, L2 visual learners were still able to comprehend at comparable depth as their captioned counterparts in a multimodal listening environment.

Aside from the comparable performances between L2 visual learners under full and no caption conditions, closer examination at the standard deviation revealed a visible variance among the visual learners' listening outcome. Such variance was also manifest in the qualitative records. Unlike the auditory learners whose qualitative data exhibited more consistency in rejecting the effect of captions, the data from the visual learners showed that they perceived captions in a positive lens: while some denounced the benefits of captions, others held a more positive mindset towards the online visual aid. Specifically, some reported that captions were (1) not distracting, (2) facilitative to their understanding of the novel words (see Winke et al., 2013), and (3) useful to resolve ambiguity which helped release their anxiety (see Vanderplank, 2010). Notably, of all the participants in this study, only the visual learners reported that having captions on made them feel less cognitively demanded. The statistical variance along with the visual learners' qualitative remarks posed a question as to whether the effect of captions was in fact, "selective" to L2 visual learners with other intra-learner factors.

5.2 RQ 2: Does working memory capacity modulate the effect of captioning?

With the aforementioned speculation in mind, the second research question illuminated that working memory capacity is an important intra-learner factor at play during multimodal input processing. Although working memory as a factor alone did not modulate the effect of captioning, quantitative data revealed that there is a significant interaction between working memory, L2 learners' modality predilection, and different caption modes. Specifically, working memory significantly determined the role of learners' preferred modality in their listening performance under different caption conditions (i.e., full and no caption).

The deciding role working memory played indicated its prominent impact on comprehension in a multimodal L2 video viewing environment—a finding that corroborated with Miyake and Friedman (1998). Such an impact lent support to the hypothesis of this study, which stemmed largely from Mayer's CTML. Based on Mayer's cognitive theory, the degree to which learners benefit from multimodal input depends on how proficient they utilize the incoming sources of information. Since listening is multimodal and transient in nature, L2 learners have to be efficient in their online input processing. It can then be hypothesized that those who are less automated in storing and processing information simultaneously may find it difficult to utilize multimodal input—a phenomenon that may undermine the effect of captions which will be discussed in the ensuing paragraph.

5.2.1 Effects of caption modes on L2 learners with lower working memory capacity

As hypothesized, the effect of captions was absent from the listening performance obtained from the visual and the auditory L2 learners with *lower working memory capacity*. In fact, neither was the presence of captions facilitative nor debilitative on both types of L2 learners with lesser cognitive capacity. Such phenomenon resonates with Mayer's assumptions in CTML, in which he claimed that although multimedia-enhanced learning has the potency to enhance L2 learning outcomes, its promising effect is conditioned to the limited capacity of working memory (Mayer, 2001, 2005a).

With limited capacity to simultaneously store and process ongoing information during video watching (Baddeley, 2003), those with lower working memory capacity may struggle to execute multimodal input processing. They may not be able to effectively and actively allocate attention to utilize enriched input while watching the video. With such cognitive capacity at hand, it is reasonable to postulate that the demand from having to attend to various input in a multimodal learning environment might have exceeded their cognitive capacity, as the presence of captions made little difference to either visual or auditory L2 learners' listening comprehension. Consequently, the multimodal nature of the video—irrespective of caption conditions and modality predilections—may not impose significant difference to those with lesser working memory, which in turn, did not benefit their listening comprehension outcome.

5.2.2 Effects of caption modes on L2 learners with higher working memory capacity

Contrary to the findings on lower working memory L2 learners, the interaction between the effect of captioning and modality predilections became salient in the listening performances of L2 learners with higher working memory capacity. Both auditory and visual learners with higher working memory performed the best under their most preferable input combinations. In other words, the visual learners significantly benefited from the captioned video; while the auditory learners performed most ideally without captions. The fact that the effect of captions was determined by their idiosyncratic modality preference again, reinforced Mayer's active processing assumption in CTML: learners tend to rely on their most preferred input for the most desirable learning outcomes (Mayer, 2001, 2005a). This match-making phenomenon generated the most optimal results for both types of learners with higher (i.e., average and/or above average) working memory capacity in this study, which again verified the hypothesis proposed in Oxford (2003). Specifically, the quantitative results obtained from the auditory learners with higher working memory capacity concurred with the findings discussed in the earlier paragraphs, as the presence of captions may have disrupted their listening comprehension, which was evidenced in their listening scores.

As for their visual counterparts, full textual support led to a more advantageous listening outcome—a finding contrary to what was observed in the auditory learners. Such effect may be attributed to them reporting paying more attention to their most preferred sources of input, visualized text, rather than the aural stream—a match that generated the best multimodal input processing scenario for them. Aside from the advantage for immersing in their most preferred processing environment, having higher attentional control could also explain the visual learners' achievement. Based on the qualitative data, some of them reported that captions were not distracting or cognitively demanding. They also self-assessed as more capable of attending to the multimodal input, which allowed them to dexterously utilize any given input during online processing (Colflesh & Conway, 2007). More control in processing implies more use of captions to enhance form-meaning mapping that aids listening comprehension (Chai & Erlam, 2008; Danan, 2004; Markham & Peter, 2003; Sydorenko, 2010; Taylor, 2005;

Vanderplank, 2010; Winke et al., 2010). Therefore, it is reasonable to suggest that higher working memory enabled the visual learners to efficiently use captions to their advantage, and consequently generated deeper input processing that led to a more optimal listening outcome.

CHAPTER 6

CONCLUSIONS

6.1 Conclusion and implications

This study showed that the effect of captions was significantly affected by intermediate-to-high intermediate L2 learners' input processing profiles, namely modality preference and working memory capacity. Specifically, among the two variables investigated in this study, working memory capacity determined the effect of modality preference in regards to the extent to which L2 learners benefited from captions. While the effect of captions was found absent in the lower working memory L2 learners regardless of their modality preferences, it was evident in their higher working memory counterparts. Among L2 learners who have higher working memory capacity, captions were significantly detrimental to the auditory learners and notably facilitative to the visual learners' listening comprehension. Consequently, the findings from this study jointly reinforced the importance to consider individual factors—highly related to L2 multimodal input processing—when providing multimedia support in differentiated teaching/learning settings.

In light of the findings of this study, the following specific suggestions are offered regarding the use of captions as a potential listening comprehension aid for L2 learners who are intermediate-to-high intermediate proficiency level. Frist, language instructors should be sensitive to L2 learners' input processing profiles when using captions to enhance their understanding of the audiovisual input. As evidenced in this study, not all L2 learners benefited from captions equally. Thus, without considering L2 learners' input processing profiles, the effect of captions in L2 listening comprehension may not be maximized according to different individual needs.

6.1.1 Differentiated instructions for optimal implementations of captions in multimodal listening

As noted earlier, intermediate-to-high intermediate L2 learners benefited from captions the most when their input processing factors were taken into consideration. Since individual factors were found highly crucial in determining the effect of captions, language instructors are encouraged to use the assessments of modality preference and working memory capacity (i.e., CRT and Reading Span Test) to diagnose L2 learners prior to exposing them to multimodal learning environment.

As evidenced in this study, those with average or above-average working memory capacity should be more aware of their real-time modality preference when using captioned video. In this group of L2 learners, captions are not recommended for those who prefer to listen, but are highly encouraged for those who are visually inclined. For the auditory L2 learners with better working memory capacity, videos with captions may not be a desirable listening environment for the presence of captions may impose extra load to their information processing. They may also be advised that processing preference plays a huge role in affecting their general understanding when receiving multimodal input. Aligning with this principle, their visual counterparts may be informed that due to having more cognitive/attentional capacity in utilizing visual input, they are more likely to make better use of captions to assist their multimodal comprehension.

As for L2 learners with below-average working memory capacity, they should be informed that their modality preference does not affect the extent to which they benefit from captions. Also, captions may not provide overt assistance to their listening comprehension. Nonetheless, qualitative data in this study showed that there is an attitudinal discrepancy between the visual and the auditory learners' listening preference. With the overall results suggesting that the visual learners are more prone-caption and auditory learners are more resistant towards captions, this may be a referential indication for lower working memory L2 learners when exposed to multimodal listening environment, which still warrants more empirical evidence.

The above differentiated instruction may be more suitable for listening practice beyond the classroom settings, for L2 learners, in particular EFL learners, tend to use videos to extend their exposure to native oral input after class. However, this does not mean that such a differentiated instruction has no practical place in the classroom setting. As shown in the participants' qualitative comments, visual and auditory learners use their focal attention for different targets in multimodal input. With this differences in mind, instructors can make good use of this point in a dictogloss dictation activity. Specifically, language instructors can pair up visual and auditory learners, inviting them to view the video materials first without captions and then immediately review the same video with captions; in both viewings, all learners have to take notes. Note that initial video viewing should not include captions so that L2 learners－irrespective of their modality preference and working memory capacity－focus only on the essence of the content without being cognitively loaded. Next, both visual learners and auditory learners would view the video again, but this time with captions. In this review, the visual learners can either focus on the captions or resprocess other visual content in the video, while auditory learners, having had the chance to process the meaning of the video once without the distraction of captions, might not find the captions distracting or disruptive－a pedagogical tip to offload the constraint of working memory capacity for auditory learners. the distinctive processing preferences of the visual and auditory learners may make their notes complementary to each other, thereby optimizing the effect of the dictogloss activity.

6.1.2 The implementation of CRT and RST for pedagogical purposes

Importantly, both CRT and RST—instruments that were originally designed as research tools—can be adapted and used as pedagogical tools by language instructors, L2 learners, or even online language learning developers. In the present study, both CRT and RST were crucial assessments that were used to determine L2 learners' real-time preferred modality and working memory capacity, respectively. Specifically, CRT, which takes the form of a multiple-choice comprehension test (a common assessment tool in the language classroom), can be easily adapted for diagnostic purposes (i.e., assessing L2 learners' preferred modality) as in the case of this study, Leveridge and Yang's (2013a, 2013b) studies and Liu and Todd's (2014) study.

The RST, on the other hand, is a cognitive assessment that can be easily adapted into a pedagogical/diagnostic tool for language teachers. First, the making of RST is not technically demanding. It can be made and presented on Power Point Presentations. Also, RST can be implemented on an individual or a group of participants. For L2 individuals with self-learning in mind, they can complete the assessment online on their own; for language instructors and researchers, RST can be implemented in a classroom or a laboratory setting.

Another added benefit of CRT and RST is that it can be incorporated into popular online language learning websites for individualized learning purposes. *VoiceTube*, for example, is a popular online learning platform that offers the assistance of captions. In light of the major findings of this study, software developers from these websites can consider creating a series of assessments, with an aim to personalize multimodal learning. This way, L2 learners are able to not only understand their input processing profiles, but also find the most optimal learning path to maximize their listening comprehension.

6.2 Future suggestions and limitations

Despite the above insights, several limitations are identified for future references. First, conclusions were drawn from small sample size in each learning conditions. For researchers interested in replicating this study, a larger sample should be used to generalize the findings in this study. Second, although the number of trials in the RST in this study is adequate (Friedman & Miyake, 2004), it can be increased to maximize the reliability of the span scores. Nevertheless, it should be noted that the length of RST should be tailored to avoid causing attention fatigue from the participants. Third, because this study focused on listening comprehension, the findings may not shed light on other long-term aspects of learning, such as listening acquisition. Whether individual factors also modulate the effect of captions on different aspects of L2 listening acquisition still warrants more future investigations.

References

Al-Shehri, S., & Gitsaki, C. (2010). Online reading: A preliminary study of the impact of integrated and split-attention formats on L2 students' cognitive load. *ReCALL, 22*(3): 356–375.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47–89). San Diego, CA: Academic Press.

Bird, S.A., & Williams, J.N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23*(4), 509–533.

Bloomfield, A., Wayland, S. C., Rhoades, E., Blodgett, A., Linck, J., & Ross, S. (2010). *What makes listening difficult? Factors affecting second language listening comprehension*. College Park, MD: University of Maryland, Center for Advanced Study of Language.

Chai, J., & Erlam, R. (2008). The effect and the influence of the use of video and captions on second language learning. *New Zealand Studies in Applied Linguistics, 14*(2), 25–44.

Chang, C. C., Tseng, K. H., & Tseng, J. S. (2011). Is single or dual channel with different English proficiencies better for English listening comprehension cognitive load and attitude in ubiquitous learning environment? *Computers & Education, 57*(4), 2313–2321.

Colflesh, G. J., & Conway, A. R. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin & Review, 14*(4), 699–703.

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta, 49*(1): 67–77.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior, 19*, 450–466.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 972–992.

Engle, R.W., Kane, M.J., & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and* *executive control* (pp. 102–134). New York: Cambridge University Press.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory & Language, 51*, 136–158.

Garza, T. J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals, 24*(3), 239–258.

Gebhard, J. (2000). *Teaching English as a foreign or second language: A teacher self - development and methodology Guide.* United States of America: The University of Michigan Press.

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28, 55–75.

Gruba, P. (2004). Understanding digitized second language videotext. *Computer Assisted Language Learning, 17*(1), 51–82.

Guichon, N., & Mclornan, S. (2008). The effects of multimodality on L2 learners: Implications for CALL resource design. *System, 36*(1), 85–93.

Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology, 21*(6), 713–729.

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

Huang, H.C., & Eskey, D. E. (1999). The effects of closed-captioned television on the listening comprehension of intermediate English as a second language (ESL) students. *Journal of Educational Technology Systems, 28*(1), 75–96.

Jong, T. D. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105–134.

Leveridge, A. N. (2015). The removal of target language captioning supports. In Turel, V. (Ed), *Intelligent design of interactive multimedia listening software* (pp. 75–101). Hershey: Information Science Reference.

Leveridge, A. N., & Yang, J. C. (2012). Effect of medium: A conceptual framework for the removal of supporting captions for EFL listening comprehension in multimedia instructional delivery. *In: 15th International CALL Research Conference: Proceedings*.Taiwan: Taichung.

Leveridge, A. N., & Yang, J. C. (2013a). Learner perceptions of reliance on captions in EFL multimedia listening comprehension. *Computer Assisted Language Learning, 27*(6), 545-559.

Leveridge, A. N., & Yang, J. C. (2013b). Testing learner reliance on caption supports in second language listening comprehension multimedia environments. *ReCALL, 25*(02), 199-214.

Liu, Y. T., & Todd, A. G. (2014). Dual-modality input in repeated reading for foreign language learners with different learning styles. *Foreign Language Annals, 47*(4), 684–706.

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 147–158). New York: Cambridge University Press.

Lwo, L., & Lin, M. C. T. (2012). The effects of captions in teenagers' multimedia L2 learning. *ReCALL, 24*(2), 188–208.

Markham, P. (1989). The effects of captioned television videotapes on the listening comprehension of beginning, intermediate and advanced ESL students. *Educational Technology, 29*, 38–41.

Markham, P. L., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems, 31*(3), 331–341.

Markham, P., Peter, L., & McCarthy, T. (2001). The effects of native language vs. target language captions on foreign language students' DVD video comprehension. *Foreign Language Annals, 34*(5), 439–445.

Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review, 8*(4), 357–371.

Mayer, R. E. (1999). Research-based principles for the design of instructional messages: The case of multimedia explanations. *Document Design Journal of Research and Problem Solving in Organizational Communication*, *1*(1), 7–19.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E. (2005a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.

Mayer, R. E. (2005b). Principles for managing essential processing multimedia learning: Segmenting, pretraining, and modality principles. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 169–182). New York: Cambridge University Press.

Mayer, R. E. (2010). Applying the science of learning to medical education. *Medical Education*, 44: 543–549.

Miyake, A., & Friedman, D. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. F. Healy & L. E. Bourne, Jr (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

Mirzaei, M. S., Meshgi, K., Akita, Y., & Kawahara, T. (2017). Partial and synchronized captioning: A new tool to assist learners in developing second language listening skill. *ReCALL, 29*(02), 178–199.

Moreno, R. (2006). Optimizing learning from animations by minimizing cognitive load: Cognitive and affective consequences of signaling and segmentation methods. *Applied Cognitive Psychology, 21*, 1–17.

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments. *Educational Psychology Review, 19*, 309–326.

Oxford, R. L. (2003). Language learning styles and strategies: Concepts and relationships. *IRAL, 41*(4)*,* 271–278.

Perez, M. M., Noortgate, W. V. D., & Desmet, P. (2013). Captioned video for L2 listening and vocabulary learning: A meta-analysis. *System, 41*(3), 720–739.

Pujola, J.T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL,14*(2), 235–262.

Reid, J. (1995). *Learning styles in the ESL/EFL classroom*. Boston: Heinle & Heinle.

Révész, A. (2012). Working memory and the observed effectiveness of recasts on different L2 outcome measures. *Language Learning, 62*(1), 93–132.

Scarcella, R. & Oxford, R. (1992). *The tapestry of language learning: The individual in the communicative classroom*. Boston: Heinle & Heinle.

Stewart, M. A., & Pertusa, I. (2004). Gains to language learners from viewing target language closed-captioned films. *Foreign Language Annals, 37*(3), 438–442.

Sun, Y., & Dong, Q. (2004). An experiment on supporting children's English vocabulary learning in multimedia context. *Computer Assisted Language Learning, 17*(2), 131–147.

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning and Technology*, 14, 50–73.

Taylor, G. (2005). Perceived processing strategies of students watching captioned video. *Foreign Language Annals, 38*(3), 422–427.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154.

Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics, 24*, 3–25.

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching, 43*(1), 1–37.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philos Trans R Soc Lond B Biol Sci, 369*, 1–7.

Winke, P., Gass, S., & Sydorenko, T. (2010). The effects of captioning videos used for foreign language listening activities. *Language Learning & Technology, 14*(1), 65–86.

Winke, P., Gass, S., & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal, 97*(1), 254–275.

Appendix 1 Information Sheet and Consent Form

STUDY INFORMATION SHEET:

I am a Master's student at the Department of English, National Taiwan Normal University (NTNU) interested in second language listening and speaking. I would like to invite you to participate in a research study that examines how different online processing profiles modulate the effect of captioning on second language listening comprehension.

If you decide to participate, I will ask you to take part in two 1-hour sessions in a computer laboratory at NTNU. In the 1st session, you will be asked to take a caption reliance test and a working memory test. In the 2nd session, you will first be asked to view a caption/no caption video, and then take a listening comprehension test. A questionnaire will be administered to you at the end of the 2nd session.

In return for your participation, I am able to offer you *NT 100* at the end of the 2nd session. As soon as the data analysis is completed, I will also share the overall results of the study with you.

Any data obtained from you will be kept securely. At every stage of the project and beyond, your name will remain confidential. Your identity will be anonymized by the use of a unique identifier. The overall results of the study will be used for my maters research project and not be shared with others. The results will also be presented at professional conferences and in research publications.

You are free to withdraw from the study at any time without reason and without any impact on you. If you decide to withdraw, any data collected from you will be destroyed. If you have any queries about the study, please feel free to contact Emily Kam at emilyfenkam@gmail.com, 0988103504.

I would be very grateful if you would agree to take part!

Emily Kam
Master's student at the Department of English, NTNU.

Consent Form

		Yes	No
1.	I have read and had explained to me by Emily Kam the Information Sheet relating to this project.	□	□
2.	I have had explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements for my participation as described in the Information Sheet.	□	□
3.	I understand that my participation is entirely voluntary and that I have the right to withdraw from the project any time.	□	□
4.	I agree with the contents of this Consent Form and have received the accompanying Information Sheet.	□	□

Name:

Signed:

Date:

Appendix 2 Listening Comprehension Questions

Name___________________

In this section, you will answer some questions about the video you just watched. Please answer the questions based on what is stated or implied by the speaker.

1. What is the main purpose of this talk?
   (A) To introduce an innovative marketing strategy.
   (B) To unfold the secret of successful leadership.
   (C) To explain an important biological concept.
   (D) To criticize on the common way of thinking.

2. Why did the speaker start his talk with a series of questions?
   (A) To express doubts.
   (B) To conduct a survey.
   (C) To capture attention.
   (D) To obtain feedback.

3. Why did the speaker begin the talk with three well-known examples?
   (A) To raise awareness on the mentioned issues.
   (B) To highlight the impact of the historical incidents.
   (C) To introduce prominent cases in the high-tech industry.
   (D) To stress the commonality of the successful cases.

4. Which of the following serves as the negative example of the "golden circle"?
   (A) Martin Luther King.
   (B) Gateway Incorporation.
   (C) The Wright Brothers.
   (D) Apple Incorporation.

5. According to the author, all great leaders and organizations are likely to communicate from "the inside out". What does this quoted term mean?
   (A) To enlighten others with their belief.
   (B) To inspire others with their experience.
   (C) To challenge others with their knowledge.
   (D) To encourage others with their wisdom.

6. Watch the following excerpt again and answer a question.

"If Apple were like everyone else, a marketing message from them might sound like this: "We make great computers. They're beautifully designed, simple to use and user friendly. Want to buy one?" "Meh."

What is the author's attitude toward this marketing message?

(A) Fascinated
(B) Disappointed
(C) Intrigued
(D) Uninterested

7. According to biology, what happens when we communicate from the "inside out"?

(A) We activate the brain area that monitors emotions.
(B) We disconnect the brain area that controls behaviors.
(C) We stimulate the brain area that generates intuitions.
(D) We damage the brain area that manages decision-making.

8. Watch the following excerpt again and then answer a question.

"This is where gut decisions come from. Sometimes you can give somebody all the facts and figures, and they say, "I know what all the facts and details say, but it just doesn't feel right.""

What does "gut decision" mean?

(A) To make decisions based on rationales.
(B) To make decisions based on intuitions.
(C) To make decisions based on suggestions.
(D) To make decisions based on tendencies.

9. Which of the following contributes to the fall of Samuel Pierpont Langley?

(A) Insufficient fundings.
(B) Poor collaborations.
(C) Wrong motivations.
(D) Bad market conditions.

10. What is the biggest difference between the people working for the Wright Brothers and those working for Samuel Pierpont Langley?

(A) The former work for a belief while the latter work for money.
(B) The former work for money while the latter work for a belief.
(C) Both of them work for money.
(D) Both of them work for a belief.

11. What is the speaker's attitude toward the quitting of Samuel Pierpont Langley on the day when the Wright Brothers took flight?
(A) Neutral.
(B) Defensive.
(C) Encouraging.
(D) Disapproved.

12. Which of the following best shows the organization of the talk?
(A) Problem → Examples → Solution
(B) Proposal → Arguments → Conclusion
(C) Introduction → Comparison → Contrast
(D) Introduction → Examples → Conclusion

13. What is the main message that the author wants to reinforce?
(A) Humans are attracted to large profits.
(B) Humans are devoted to effective leaders.
(C) Humans are influenced by what others think.
(D) Humans are driven by a sense of meaning.

14. What is the best title of this talk?
(A) How inspiring speakers covey ideas.
(B) How famous inventors achieve success.
(C) How great leaders inspire actions.
(D) How leading companies train employees.

15. If you were to give advice to a friend who wants to become a successful entrepreneur, which of the following best corresponds with the speaker's idea?
(A) See every detail of your business.
(B) Being comfortable is your enemy.
(C) You are who you associate with.
(D) Bring your purpose into action.

Appendix 3 Questionnaires

(Questionnaire for no caption condition)

Part 1 General experience of watching captioned videos

Below are a number of statements regarding your *general* experience of watching captioned videos. Please read each statement carefully and indicate to what extent you agree or disagree with it on a five-point Likert scale (1=strongly disagree; 5=strongly agree).

	Statements	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	I use captioned videos to learn English.	1	2	3	4	5
2	I use captioned videos to enhance my L2 listening skills.	1	2	3	4	5
3	I think captions help me understand the audio content.	1	2	3	4	5
4	I find captions helpful for learning English.	1	2	3	4	5
5	I think the captions in videos are distracting.	1	2	3	4	5
6	I always attend to the captions while watching videos.	1	2	3	4	5
7	I focus more on the audio than the captions while watching videos.	1	2	3	4	5
8	I focus more on the captions than the audio while watching videos.	1	2	3	4	5

Part 2 Perception of watching the (captioned) video

Below are a number of statements to probe your video watching experience *in this study*. Please read each statement carefully and indicate to what extent you agree or disagree with it on a five-point Likert scale (1=strongly disagree; 5=strongly agree).

	Statements	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	I felt the video was interesting.	1	2	3	4	5
2	I had some background knowledge about the video topic.	1	2	3	4	5
3	I could understand the gist of the video.	1	2	3	4	5
4	I found it difficult to attend to all input (i.e., audio, captions, visual) in the video.	1	2	3	4	5

If you have any comments to the researcher, please provide them in the space below (You can use either English or Chinese to answer this question):

____

____

(questionnaire for full caption condition)

<u>Part 1 General experience of watching captioned videos</u>

Below are a number of statements regarding your *general* experience of watching captioned videos. Please read each statement carefully and indicate to what extent you agree or disagree with it on a five-point Likert scale (1=strongly disagree; 5=strongly agree).

	Statements	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	I use captioned videos to learn English.	1	2	3	4	5
2	I use captioned videos to enhance my L2 listening skills.	1	2	3	4	5
3	I think captions help me understand the audio content.	1	2	3	4	5
4	I find captions helpful for learning English.	1	2	3	4	5
5	I think the captions in videos are distracting.	1	2	3	4	5
6	I always attend to the captions while watching videos.	1	2	3	4	5
7	I focus more on the audio than the captions while watching videos.	1	2	3	4	5
8	I focus more on the captions than the audio while watching videos.	1	2	3	4	5

Part 2 Perception of watching the (captioned) video

Below are a number of statements to probe your video watching experience *in this study*. Please read each statement carefully and indicate to what extent you agree or disagree with it on a five-point Likert scale (1=strongly disagree; 5=strongly agree).

	Statements	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
1	I felt the video was interesting.	1	2	3	4	5
2	I had some background knowledge about the video topic.	1	2	3	4	5
3	I could understand the gist of the video.	1	2	3	4	5
4	I found it difficult to attend to all input (i.e., audio, captions, visual) in the video.	1	2	3	4	5
5	I thought the captions helped me understand the audio content.	1	2	3	4	5
6	I found the captions helpful for learning English.	1	2	3	4	5
7	I found the captions in the video were distracting.	1	2	3	4	5
8	I always attended to the captions while watching the video.	1	2	3	4	5
9	I focused more on the audio than the captions while watching the video.	1	2	3	4	5
10	I focused more on the captions than the audio while watching the video.	1	2	3	4	5