INFLUENCE OF MULTIMEDIA HINTS ON CONCEPTUAL PHYSICS PROBLEM
SOLVING AND VISUAL ATTENTION

by

XIAN WU

B.S., Wuhan University, 2007
M.S., University of New Hampshire, 2012

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Abstract

Previous research has showed that visual cues can improve learners' problem solving performance on conceptual physics tasks. In this study we investigated the influence of multimedia hints that included visual, textual, and audio modalities, and all possible combinations thereof, on students' problem solving performance and visual attention. The participants (N = 162) were recruited from conceptual physics classes for this study. Each of them participated in an individual interview, which contained four task sets. Each set contained one initial task, six training tasks, one near transfer task and one far transfer task. We used a 2 (visual hint/no visual hint) × 2 (text hint/no text hint) × 2 (audio hint/no audio hint) between participant quasi-experimental design. Participants were randomly assigned into one of the eight conditions and were provided hints for training tasks, corresponding to the assigned condition. Our results showed that problem solving performance on the training tasks was affected by hint modality. Unlike what was predicted by Mayer's modality principle, we found evidence of a reverse modality effect, in which text hints helped participants solve the physics tasks better than audio hints. Then we studied students' visual attention as they solved these physics tasks. We found the participants preferentially attended to visual hints over text hints when they were presented simultaneously. This effect was unaffected by the inclusion of audio hints. Text hints also imposed less cognitive load than audio hints, as measured by fixation durations. And presenting visual hints caused more cognitive load while fixating expert-like interest areas than during the time intervals before and after hints. A theoretical model is proposed to explain both problem solving performance and visual attention. According to the model, because visual hints integrated the functions of selection, organization, and integration, this caused a relatively heavy cognitive load yet improved problem solving performance. Furthermore, text hints were a better

resource for complex linguistic information than transient audio hints. We also discuss

limitations of the current study, which may have led to results contrary to Mayer's modality

principle in some respects, but consistent with it in others.

INFLUENCE OF MULTIMEDIA HINTS ON CONCEPTUAL PHYSICS PROBLEM
SOLVING AND VISUAL ATTENTION

by

XIAN WU

B.S., Wuhan University, 2007
M.S., University of New Hampshire, 2012

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Physics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Brett DePaola

Copyright

XIAN WU

2016

Abstract

Previous research has showed that visual cues can improve learners' problem solving performance on conceptual physics tasks. In this study we investigated the influence of multimedia hints that included visual, textual, and audio modalities, and all possible combinations thereof, on students' problem solving performance and visual attention. The participants (N = 162) were recruited from conceptual physics classes for this study. Each of them participated in an individual interview, which contained four task sets. Each set contained one initial task, six training tasks, one near transfer task and one far transfer task. We used a 2 (visual hint/no visual hint) × 2 (text hint/no text hint) × 2 (audio hint/no audio hint) between participant quasi-experimental design. Participants were randomly assigned into one of the eight conditions and were provided hints for training tasks, corresponding to the assigned condition. Our results showed that problem solving performance on the training tasks was affected by hint modality. Unlike what was predicted by Mayer's modality principle, we found evidence of a reverse modality effect, in which text hints helped participants solve the physics tasks better than audio hints. Then we studied students' visual attention as they solved these physics tasks. We found the participants preferentially attended to visual hints over text hints when they were presented simultaneously. This effect was unaffected by the inclusion of audio hints. Text hints also imposed less cognitive load than audio hints, as measured by fixation durations. And presenting visual hints caused more cognitive load while fixating expert-like interest areas than during the time intervals before and after hints. A theoretical model is proposed to explain both problem solving performance and visual attention. According to the model, because visual hints integrated the functions of selection, organization, and integration, this caused a relatively heavy cognitive load yet improved problem solving performance. Furthermore, text hints were a better

resource for complex linguistic information than transient audio hints. We also discuss

limitations of the current study, which may have led to results contrary to Mayer's modality

principle in some respects, but consistent with it in others.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Dedication

To my parents.

In memory of my grandparents:

Yishen Wu (August 25, 1919 – May 17, 2001)

Yahui Zhang (August 30, 1921 – August 25, 2006)

Youxin He (May 14, 1924 – March 12, 2013)

Xiaofang Peng (January 6, 1925 – September 7, 2001)

Chapter 1 - Introduction

Physics Education Research (PER) has a long tradition of research on understanding student thinking (McDermott, 2001; Redish, 2004; Docktor and Mestre, 2014). Since the beginning, PER as a research domain, has expanded and developed into several different subfields. Among those subfields, physics problem solving is one of the most important subfields (Hsu, *et. al*., 2004; Adams and Wieman, 2015).

As a starting point, the definition of what we deem as a "problem" needs to be clarified. Several different problem types have been addressed in research: end-of-chapter problems (Chi and VanLehn, 1991), authentic problems (AAAS, 2011; Gormally *et al*., 2012), complex problems (Goldenfeld and Kadanoff, 1999), context rich problems (Heller and Hollabaugh, 1992) and many others. Three aspects of physics problem solving are analyzed most frequently: cognitive processes, content knowledge, and mathematics strategies. We are interested in the cognitive aspect; especially the difficulties students have with building and shifting their mental representation of the problem during problem solving. Therefore, in the study presented in this dissertation, we adopt Jonassen's (2011) definition on problems: "question or issue that is uncertain and so must be examined and solved" (pp. 1). Due to the diversity of problem definitions, we use the term "problem" and "task" interchangeably in this dissertation.

A physics education researcher is not only a scientific researcher, but also a science educator. There are lots of efforts devoted in helping students solve physics problems. Both the content and the modality of the hint information provided to assist the learner in solving the problem could affect the effectiveness of the hint. In our everyday life, we use pictures, videos, text, and speech to communicate our ideas. How can we apply these modalities to help our students learn, and how can we assess the learning outcomes? Moreover, can we model the

interactions between learners and the information in different modalities? Also, how can we use these insights to facilitate problem solving? These are critical questions that do not yet have satisfactory answers.

Physics education researchers typically rely on observations, surveys, tests, and interviews to measure student thinking and learning outcomes. It is difficult to gain insights into students' perception and cognition using these tools. A student's brain is typically a black box to physics education researchers, and its functioning can only be inferred by the input and output. Now, with the advent of technology it is possible to directly measure the gestures, facial expressions, eye movements, and brain activity of students. In this dissertation, we discuss a study on physics problem solving with multimedia hints using eye-tracking technology. The study utilizes eye movements as an important source of insight with regard to student thinking, and in conjunction with their performance on problem solving tasks. In this chapter, we begin with the motivation for this study. Then we will discuss the research questions, research approach, and the organization of the dissertation.

Background and Motivation

The studies presented in this dissertation were motivated by eye movement studies on problem solving (Knoblich, *et al*., 2001; Grant and Spivey, 2003; Thomas and Lleras, 2007). Knoblich *et al*. (2001) used the results of eye movement data analyses to support Ohlsson's (1992) problem solving model: Representational Change Theory (RCT). The studies conducted by Grant and Spivey (2003) revealed the connection between eye movement and cognition during problem solving with pictorial information. Solvers' eye movements indicated how solvers considered the problem. Further, Thomas and Lleras found (2007) guiding solvers' eyes to attend the area of diagram where directly related to correct answer could improve problem-

2

solving performance. These studies opened a new window for researchers who are interested in problem solving, and they inspired our group's eye movement studies on physics problem solving on many aspects. First, we adopted Ohlsson's RCT as the theoretical framework of problem solving. Second, the eye-tracking technology was introduced to grant us insights concerning the process of solving physics problems. These two aspects will be elaborated on in the later sections of this dissertation. We designed a set of physics tasks with pictorial information. There were two distinct areas on each picture. The first, was associated with a well-documented incorrect answer, and the second, was associated with the correct answer. We measured students' problem solving performance and eye movements. In the previous study (Madsen, *et. al*., 2012; 2013), we found that student's overt visual attention was correlated with student's prior knowledge. In other words, the correct solvers attended to the correct answer areas more than the incorrect solvers did, and the incorrect solvers attended to the incorrect answer areas more than the correct solvers did. This raised the question as to whether or not guiding students to attend toward the correct answer areas would help them solve the problem correctly. The following study (Rouinfar, 2014) demonstrated that this was indeed the case. The students with both visual cues and feedback solve the training and transfer problems significantly better than all the other conditions. Next, a theoretical effort was devoted by Agra's study (2015). The measure of confidence and delayed transfer task solving performance have been adopted along with the eye movement data to explore the mechanism of mental representation change in physics problem solving.

All of the previous studies used visual cues to facilitate changes in student thinking while solving problems. However, visual cues are not the only type of help information. Compared to visual cues, linguistic hints are more explicit. Therefore, we would like to include the linguistic

hints in our study. We assess the effectiveness of all three modalities of hints: visual, audio, and text, and all possible combinations. Presenting hints via multiple modalities allows us to step into the world of multimedia. A rich body of research on multimedia instruction and multimedia learning has provided both experimental evidence and theoretical framework for understanding learning on multimedia instruction and guiding multimedia instruction design. Mayer's Cognitive Theory of Multimedia Learning (CTML, Mayer, 2002), is the theory that has been continuously tested in many different disciplinary domains and contexts, and has demonstrated positive results in most of the cases.

Mayer's CTML has two major implications. First, it provides theoretical explanation in regard to the cognitive process of multimedia learning. Second, it generates a set of principles to guide multimedia instructional design. Both aspects shed light on the studies presented in this dissertation. The theory aspect of CTML provides us the theoretical foundation for understanding perception of multimedia material. The design principles of CTML polish our multimedia material design (not sure what this sentence means). Some of these principles are used to help generate hypotheses toward studying the performance of problem solving. We investigate if (1) linguistic hints are a superior resource of help information than visual hints in physics problem solving, (2) linguistic hints could assist visual hints in being more effective at helping physics problem solving, (3) sending the identical linguistic information via different modalities would affect physics problem solving. This was the motivation for Research Question 1 of this study.

Further, we want to know how students attend to multi-modality hints. This is an exploration of students' eye movements. As reviewed in the previous paragraph, the connection between cognitive processes and visual attention is evident in studies of the eye movements

during problem solving. Our group's previous studies also relied on the evidences from eye movement study to reveal part of the cognitive process of physics problem solving. Building upon these results, we hope to conduct a study to understand the connection between perceiving hints and visual attention. This motivates Research Question 2 of this study.

Research Question 1 focuses on "what" modalities of hints facilitate students to solve conceptual physics tasks with pictorial representation. Research Question 2 focuses on "why" certain types of hint modalities are more effective than others at facilitating problem solving. Finally, Research Question 3 addresses "how" different modalities of hints facilitate problem solving. With the results of two previous research questions, we model students' cognitive processes of conceptual physics problem solving with multimedia hints. The model could be used to explain physics problem solving and guide the design of physics instructional materials. This motivates Research Question 3.

Research Questions

The overarching question of this study is to explore the effect of multimedia hints on physics problem solving. Specifically, we want to answer three research questions.

1. How does the combination of visual, text, and audio hints affect students' performance on solving introductory conceptual physics tasks with graphic representation?

2. How does hint modality influence students' visual attention?

(a) How does hint modality affect the cognitive load during problem solving?

(b) How does simultaneously presenting multi-modality hints split students' visual attention during problem solving?

3.   What theoretical model can explain how multi-modality hints influence both students' problem solving performance and their visual attention while solving conceptual physics tasks?

Organization of Dissertation

The major thrust of this dissertation, an eye-tracking study of physics problem solving with multimedia hints, contains two sub-studies. Chapter 2 discusses the effect of multimedia hints on problem solving performance. Participant problem solving performances on initial tasks, training tasks, and transfer tasks were analyzed. We found hint modality affected participant performance on solving training tasks. This finding motivated the second sub-study, which is described in Chapter 3. This sub-study focused on participant eye movement data. Two categories of eye movement data were included: one is the domain relative ratio, and another is the mean fixation duration. The domain relative ratio is the measure of visual attention distribution, and the mean fixation duration indicates the real-time mental effort. We found evidence from our eye movement study to support our finding from the study of participant performance. Chapter 4 assembles the experimental evidence from two sub-studies and the theoretical frameworks of multimedia learning, problem solving, and working memory together to construct a theoretical model of physics problem solving with multimedia hints. This model provides a coherent explanation of results of both the problem solving performance and eye movement data, and generates a set of hypotheses to guide future studies. Chapter 5 provides a summary of the results of the previous chapters. It shows how research questions have been addressed by these results and discusses the possible implications of the study, and offers suggestions for future work.

Chapter 2 - Multimedia Hints Improve Problem Solving Performance

Introduction

Problem solving is regarded as one of the critical skills of learning physics and has been investigated by the PER community for decades (Hsu, *et al*., 2004; Adams and Wieman, 2015). There are several studies (Chi *et al*., 1981; de Jong and Ferguson-Hessler, 1986; Hardiman *et al*., 1989) showing that it is very difficult for novices to solve physics problems. Physics problems often have pictorial information. Therefore, it is important for us as physics educators and PER researchers to understand how to facilitate learners to solve physics problems that include pictorial representations.

Previously, we studied the effects of visual cueing and feedback on physics problem solving (Rouinfar, 2014). That study included 90 participants that were randomly assigned in a 2 (cue/no cue) × 2 (feedback/no feedback) design. Each participant was asked to solve tasks from four task sets, whose order was also randomized. Each task set has one initial task, six training tasks presented to participants in a random order, and one transfer task. The participants who were in a cued condition were provided a visual cue as they solved each training task. The cues were shapes or lines superimposed on the diagram that would direct the task solvers' attention to relevant features in the task in a way that would facilitate them to solve tasks. Participants in the feedback condition were told whether their responses (answers and reasoning considered together) were correct or incorrect after they responded to an interviewer. Results showed that the participants who received both visual hints and feedback (cue + feedback condition), had a significantly higher performance than the other conditions on the training tasks as well as the transfer tasks across all task sets.

Rouinfar's study (2014) had two limitations, both pertaining to the ecological validity of the study. First, the feedback was provided by a human interviewer. This is inconsistent with hints administered in the context of computer-aided instruction, where hints would be provided automatically based on learners' responses. Second, the study did not explore the use of other modalities, in addition to the visual modality, which would be relevant in multimedia-based computer aided instruction. While the first limitation is currently being investigated in a different study, the study described in this dissertation addresses the second limitation.

In this study, we focus on the effect of multimedia hints. Rather than purely visual cues used in the Rouinfar (2014) study, we tested the effect of hints provided using textual and auditory modalities. The existing theoretical guidelines for multimedia instruction design are adopted from Mayer's (2002) CTML and Wickens' (2002) Multiple Resources Theory (MRT). These two theories draw a map describing how learners perceive and understand external information that is presented in visual and auditory modalities. CTML proposes principles that could potentially guide the development of content for computer-aided instruction, especially with regard to the appropriate modalities and combinations thereof. MRT predicts task performance in the dual-task scenario, i.e., the scenario that humans are engaging multiple tasks simultaneously. There have been numerous studies that have tested each of the principles of Mayer's CTML and Wickens' MRT. These two theories have also been used to guide the development of online learning environments to facilitate physics learning (Sadaghiani, 2011; Salim, *et al*., 2012) and the optimization of human-machine interaction (Stork, *et al*., 2008). But, to date, there are no published studies in which these theories have been tested in the context of physics problem solving. Therefore, our current research investigates the applicability of CTML and MRT for helping learners solve conceptual physics problems.

Theoretical Background

Generally speaking, the design of computer-assisted instruction reflects the designer's understanding of the user's learning process. The underlying hypothesis of studies around computer-assisted instruction design is that the content and representation of the instruction should fit the user's learning mechanism (Leahy *et al*., 2003; Kirschner *et al*., 2006; Smith, 2006; Pashler *et al.*, 2007). According to Mayer's CTML (2002), there are three theoretical assumptions that underpin human interaction with multimedia instruction. Below, we describe each assumption in detail.

The first assumption in CTML is the *dual-channel assumption,* which states that leaners have two separate systems to process pictorial and verbal information. In his theory, Mayer (2002) adopted the *sensory-modalities* approach to differentiate the visual and auditory channels. This approach categorized the human information processing system into two channels according to the modality of sensory input, i.e., vision and hearing. The two channels, therefore, are the visual channel and the auditory channel. Material presented to learners' eyes, such as pictures, animations, and on-screen text, are processed through the visual channel, and material presented to the learner's ears, such as narration, music and other sounds, are processed through the auditory channel. The second assumption is the *limited capacity assumption*, which assumes that the capability of processing material for each channel is limited. If the amount of material presented to one channel exceeds this channel's capability, the material could not be processed through this channel appropriately. This assumption is consistent with Sweller's cognitive load theory (Sweller, 1988; Chandler and Sweller, 1991; Sweller, 1999). The third assumption is *active-processing assumption*. It is a fundamental assumption that emphasizes the necessity of cognitive processing in learning.

Wickens' MRT (2002) provides another perspective of the interaction between learners and multimedia instruction. While CTML considers the modality (visual and audio) of the incoming information, MRT considers both how the information is coded (using spatial and verbal resources) as well as the modality (visual and audio) of presentation. Moreover, MRT also considers the stages (perception, cognition, and responding) of processing. While information received from the two modalities is separated in the perception stage, the distinctions between the two modalities do not exist in the cognition and responding stages of processing.

Wickens' MRT proposes an explanation of how humans process information while simultaneously engaging with multiple tasks. The main idea of MRT is that performance on tasks is negatively impacted if the multiple tasks share the same codes or modalities. For example, comparing talking on a hands-free phone while driving, versus texting, the latter would have a greater effect on driving ability, since texting and driving share the visual *modality*. Wickens' model provides another point of view for Mayer's *dual-channel,* and *limited capacity assumptions*.

These assumptions together serve as the foundation for generating twelve major principles in designing multimedia instruction (Mayer, 2005b). See Table 2-1, for a brief explanation of each principle. This study is focused on testing the *Multimedia Principle*, the *Signaling Principle*, the *Redundancy Principle*, and the *Modality Principle* in the context of physics problem solving. Most of the rest of the principles were used in the design of the multimedia hints that will be described in subsequent sections.

Table 2-1 Research-based principles for multimedia instruction design by Mayer (2005)

Principle	Description: People learn …
Multimedia	better when pictorial and linguistic information are presented together, rather than when linguistic information is presented alone.
Spatial Contiguity	better when corresponding pictures and words are near each other than when they are far away.
Temporal Contiguity	better when corresponding pictures and words are presented simultaneously than successively.
Coherence	better when presented pictures and words are related with each other than when they are unrelated.
Segmenting	better when instructional material is presented in segments that can be played by the user at their chosen rate.
Personalization	better when familiar language rather than formal, unfamiliar language is used.
Voice	better when the narration is in a familiar human voice, rather than in an unfamiliar machine voice
Image	the same regardless of whether or not the image of the speaker appears in multimedia instruction
Pre-Training	better when they have prior knowledge of the main concepts presented in instruction
Signaling	better when visual cues are used to highlight the organization of the material
Redundancy	better with animation and narration, rather than with animation and narration and text together.
Modality	better with auditory information, than with on-screen texts .

The *multimedia principle* was based on a study that investigated what kinds of illustrations can better improve learning from expository passages (Mayer and Gallini, 1990). Recently, the term *multimedia* has been used more broadly. It could include different

combinations of pictures and texts or pictures and auditory content. There are studies that showed presenting visual and linguistic information together could improve retention (Mayer, 2002), deepen understanding (Tabbers, *et al.*, 2004; Cuevas *et al.*, 2002) and aid in problem solving (Hoffler and Leutner, 2007). The benefit for retention can be explained by the *dual-channel* and *limited capacity assumptions* of CTML. Pictorial and linguistic information, are coded in parallel to benefit memorization of that information. This is also consistent with Wickens' MRT. Pictorial information is processed using the spatial resource, and linguistic information is processed by the verbal resource, therefore they do not interfere with each other. Consequently, the combination of visual and audio information will facilitate learning better than the combination of visual and textual information, because visual processing might be overloaded when pictorial and textual information are presented together, which can lead to competition for attention in the visual channel (Mayer, 2002). However, the reason why pictorial information also seems to help with deeper understanding and problem solving is still not clear (Butcher, 2014).

The *signaling principle* emphasizes the advantage of visual representation. Unlike the *multimedia principle*, which addressed the question of how to improve learning from text, the *signaling principle* focuses on how to improve learning from pictures. It is also known as the cueing principle. This principle is related to *active-processing* and *limited capacity* assumptions. According to the former, humans are active processors of information that always try to understand incoming information that they receive through their senses. Providing signals or cues can help the human brain select and organize the incoming information to avoid possible processing channel overflow, since each channel only has limited processing capacity. Our group's previous studies have addressed this principle in the physics problem-solving domain.

Madsen found that visual cues could draw participants' visual attention to specific regions of diagrams and improve the problem solving performance (Madsen, *et al*., 2013). This finding was replicated in Rouinfar's study with 2 (cue / no cue) × 2 (feedback / no feedback) design (Rouinfar, 2014).

Before the development of Mayer's CTML (Mayer, 2002), the *redundancy principle* for multimedia learning was used by Kalyuga *et al.* (1999). Simply put, the principle can be defined as: *less is better*. In Mayer's theory, this means that in the presence of animation, learning from text and narration together is less effective than learning from narration alone. This principle is directly connected to the *dual-channel* and *limited capacity* assumptions. On-screen text and animation presented simultaneously may force the learner to split their attention. The split attention may result in insufficient capacity to process presented information. However, when there is no on-screen text, narration and animation can be processed by the learner through the visual and auditory channels separately. The capacity of each channel will be the sum of the two, and give the learner more opportunity to attend to the presented information. The *redundancy principle* can also be explained by Wickens' MRT. Both text and narration need processing of verbal resources. Presenting information simultaneously via text and audio can cause interference in processing of verbal information.

Several studies have tested the *redundancy principle* and the results of these studies are mixed. Craig *et al*. (2002) found that the *redundancy principle* cannot be washed out when the text is presented close to an animation, and with an on-screen agent. The principle has also been verified in computer science teaching (Rias and Zaman, 2010) and music theory teaching (Aldalalah and Fong, 2010). Mayer and Johnson's study (2008) showed that learners might use brief on-screen text as guidance to enhance their learning. They suggested three theoretical

reasons for this result. First, short on-screen text presents a smaller cognitive load in the visual channel than longer on-screen text; second, presenting on-screen text close to the corresponding figure helps the learner attend to key information more effectively than on-screen text that is far from the corresponding figure; third, on-screen text is more effective with static figures than animation.

Some studies have demonstrated a *reverse redundancy effect,* namely that learning from text and narration together is more efficient than learning from narration alone. Toh *et al*. (2010) found the *reverse redundancy effect* in their study of English reading comprehension. In their study, the learners who received an instructional presentation with static pictures, narration, and synchronized on-screen text, outperformed the learners who received the instructional presentation with static pictures and narration only. They suggested that their learner-paced presentation and synchronized on-screen text, which is proximal to the relevant picture, "*cue* the learner to the learning task" (pp. 995). Instead of hindering learning, on-screen text helped the learner select, organize the presented information, and integrate it with prior knowledge to achieve active learning (Mayer, 2002). In a recent study, Yue *et al.* (2013) focused on the relationship between the format of on-screen text and the effect of redundancy and reverse redundancy with the content of an astronomy instructional presentation. They found that the short on-screen text actually could help the learner better learn the presented content. However, if the difference between on-screen text and narration was increased, such as by using synonyms to replace narrated words or changing sentence structure, the short on-screen text would be as ineffective as the full text or narration.

The *modality principle* in CTML states that narration presented simultaneously with pictorial information is easier for the user to learn from and retain than on-screen text with

14

pictorial information (Mayer, 2002; Low and Sweller, 2005). This principle can also be explained by *dual-channel* and *limited capacity assumptions*. Pictorial information is processed through the learner's visual channel. If textual information is presented to the leaner's visual sensory system simultaneously with pictorial information, it can cause huge cognitive load and may cause processing difficulties. However, if there were narrated information presented to a learner's auditory sensory system, it can be processed via the learner's auditory channel. So, narrated information is easier for a learner to attend to simultaneously with pictorial information. This is also consistent with Wickens' MRT. The interference between on-screen text and pictorial information is much larger than the interference between narration and pictorial information, since on-screen text and pictorial are all presented in the visual modality. This interference can hinder learning.

The results of studies around the *modality principle* are also mixed. Schmidt-Weigand and co-workers tested a modality effect with German university students (Schmidt-Weigand *et al.*, 2010b). They showed that students who received spoken text tended to remember the content better than students who received on-screen written text. On the other hand, Harskamp *et al.* (2007) tested the modality principle in an authentic biology classroom with secondary school students. Their results showed that the modality effect was washed out if students were allowed to control the pace of instruction, and allowed to review the material later. Specifically, students who needed less learning time performed better with pictures and narration instruction while students who needed more learning time performed better with pictures and text instruction. Harskamp *et al.* (2007) suggested that the modality effect might be valid only in a fast paced learning environment, but not necessarily in a self-paced learning environment. Stiller *et al.* (2009) probed the interaction between pacing effect and modality effect. They recruited

college students to watch instructional materials about the human eye and administered a pretest and posttest on the material. The experiment showed that the narration condition significantly outperformed the on-screen text condition for system-paced presentations, while the on-screen text condition significantly outperformed the narration condition for learner-paced presentations. There was no significant difference in posttest performance between the conditions in system-paced presentations. But, since the learners in the self-paced condition spent less time than the learners in the narration condition, the authors suggested that on-screen text with a learner-paced system might be the best for knowledge retention.

Significance of Study

In summary, the principles of Mayer's CTML have been widely used as guidelines in the creation of multimedia learning materials. However, there appears to be no clear consensus in the literature with regard to evidence in support of the *redundancy* and *modality principles*. Moreover, CTML principles have not been tested in the context of conceptual physics problem solving. With the increasing prevalence of computer tutors and multimedia learning systems in physics education, we believe that it is important for the field to examine the applicability of these principles in the design of the systems.

This study is a first small step toward a broader investigation of the applicability of the principles of multimedia learning to physics problem solving, especially the *redundancy* and *modality principles*. The scope of this study is confined to the use of multimedia hints in conceptual physics problems that have a diagram. The study builds on our previous work on visual cueing (Rouinfar, 2014) and incorporates other modalities of hints such as textual and audio hints used in conjunction with each other as well as with visual hints.

The overarching research question of this study is: *How does the modality of multimedia hints provided on conceptual physics tasks affect students' performance on these tasks and on subsequent transfer tasks on which no hints are provided?*

Method

Participants

The participants (N = 162) in this study were recruited from conceptual physics courses at a mid-western university. Most of them are sophomores and juniors. Vast majorities (> 80%) of these students were future elementary teachers, and vast majorities (> 90%) of them were female. Very few of these students (< 10%) had any prior physics class in high school and none of the students had a prior college physics class. As incentive for participation in the study, most participants were given extra credit equivalent to 1% of the course grade for participating in the one-hour long interview. There were ten participants compensated with cash. The students were recruited via email and were provided a link that they could use to sign up for the interview at a convenient time.

Materials

Each participant solved four sets of conceptual tasks in the interview. Similar to the sets used in our previous work (Rouinfar, 2014), we name each of these after the main object in the task -- Ball, Graph, Skier and Roller Coaster. Each set had one initial task, six training tasks, one near transfer task and one far transfer task. Each of the training tasks differed from the initial task only in terms of surface features, and not in terms of deep structure, in that they had the same physics concept and the same representation, only a minor change in the details of the situation. The tasks were presented to participants with multimedia hints that will be discussed in detail in the following two sections. The near transfer task was designed based on the same

17

physics concept and representation but in a different context. The far transfer task was again

based on the same physics concept and representation, but the context was substantially different

from the context of the training and near transfer tasks. See Figure 2-1 for examples of the

initial, training, near transfer and far transfer tasks. The topics relevant to the tasks were

kinematics and energy conservation, which had been covered in lecture prior to the recruitment

of students. All of the sets were used in previous studies and they showed that correct solvers

and incorrect solvers look at training tasks and near transfer tasks significantly differently when

only the visual hint was provided to participants (Rouinfar, 2014). In our current study, we

added the far transfer task following the near transfer task in each set to test whether different

hint modalities would affect learners' performance on a far transfer task differently. As a part of

our experimental design, we randomized conditions as explained in the following two sections.

We also randomized the sequence of sets, and the sequence of training tasks within each set.

Figure 2-1 An example of an initial, training, near transfer, and far transfer task (from the top to the bottom) from the Ball task set.
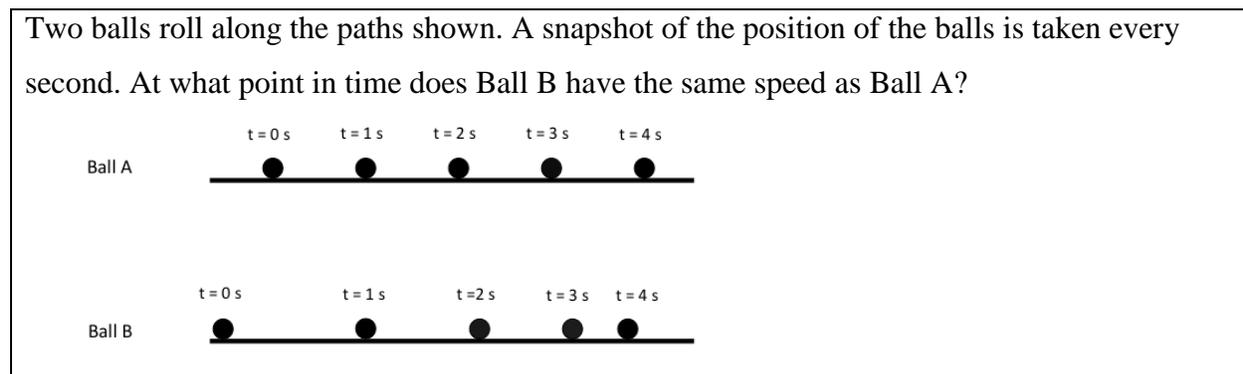
Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?

Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?

Ball A

t = 0 s t = 1 s t = 2 s   t = 3 s   t = 4 s

Ball B

t = 0 s   t = 1 s   t = 2 s   t = 3 s   t = 4 s

Ball A begins riding downward in an elevator at the same time Ball B is dropped from the roof of an adjacent building. A snapshot of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?

Ball A

t = 0 s

t = 1 s

t = 2 s

t = 3 s

t = 4 s

t = 0 s   Ball B

t = 1 s

t = 2 s

t = 3 s

t = 4 s

A runner runs along a track. The following diagram, viewed from above, shows the position of the runner at each second. At which point in time is the runner moving the fastest?
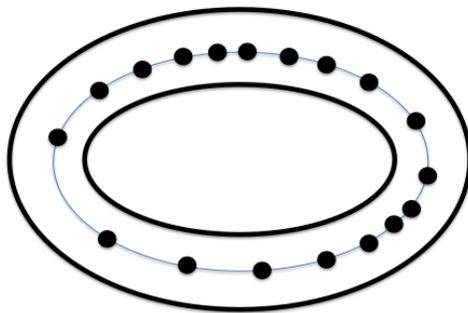
Multimedia Hint Design

The goal of this study was to test the effect of different modalities of multimedia hints on students' performance on conceptual physics tasks. Specifically, we wanted to test the effects and interactions between text hints, audio hints, and visual hints. We used a full factorial design:

19

2 (visual hint / No visual hint) × 2 (text hint / No text hint) × 2 (audio hint / No audio hint).

Participants were randomly assigned to one of the eight conditions. Participants in each

condition received hints with different modalities when they solved the training tasks, such as

text only, audio only, visual only, visual + text, visual + audio, text + audio, and visual +text +

audio, or no hints. Participants were not provided with any hints on the initial, near transfer, or

far transfer tasks in any of the conditions.

We adopted visual hints from our previous study and the more detailed explanation of

these hints can be found there (Rouinfar, 2014). In short, the visual hint for each training task is

eight-seconds long in time, highlighting the area of the task diagram that is related to the correct

answer. For example, the highlighting patterns of Figure 2-2 are an example of a visual hint.

**Figure 2-2 Examples of training tasks with visual hints and text hints superimposed from
the Skier (top), Roller Coaster, Graph, and Ball (bottom) task sets. All hints appeared on
screen for a total of eight seconds at a time.**

Compare the change in position from initial to final for two carts

A   B

Initial   Final   Initial   Final

1,3   2,4   5,7   6,8

position

1,5

Compare the slope along the curve at each moment in time

2,6

3,7

4,8

time

t = 0 s   t = 1 s   t = 2 s   t = 3 s   t = 4 s

Ball A

1   3   5   7

Compare the distance between subsequent snapshots of the two balls

t = 0 s   t = 1 s   t =2 s   t = 3 s   t = 4 s

Ball B

2   4   6   8

In our current study, we made two modifications to the visual hints used in our previous work (Rouinfar, 2014). First, in the Graph task, we presented the one red tangent line for one second each followed by the next tangent line instead of showing all of the tangent lines at the same time for eight seconds. Second, in the Skier and Roller Coaster tasks, we highlighted each height section for one second at a time instead of highlighting the whole height from the start to

21

the end for eight seconds. The main reason for those changes is that we wanted to ensure that the visual hints for every training task across all of the task sets were similar in that each included animated cues which facilitated the learner to direct their attention from one area of the task diagram to the next in a sequence that would facilitate them to make the kinds of comparisons that would solve the task. For instance, in the Graph task students would need to focus on comparing the slopes of the two graphs at the same point in time, therefore the tangent lines on each graph appeared simultaneously at each point in time to facilitate such a comparison. Similarly, for the Skier tasks the learner had to compare height lost in various sections of the slope therefore the visual hint was an animated cue that sequentially highlighted the heights of each section of the slope. Essentially, all of our task sets were asking participants to make comparisons by using the information provided by statements of tasks and presented pictures. We want to convey the idea of "making a comparison" by using the manner of sequentially highlighting aspects of the picture that students had to compare, for all of our training tasks.

We used the same idea in the design of our text hints and audio hints, to ensure that the text hint conveyed the same meaning as the corresponding visual hint. For example, we highlighted the distance between two balls for every one-second as our visual hint for the ball tasks. The text hint we designed is "Compare the distance between subsequent snapshots of the two balls." From our point of view, the word "compare" in the text hint is equivalent to the manner of sequentially highlighting elements in the diagram in the visual hint; the phrase "the distance between subsequence snapshots of the two balls" in the text hint is equivalent to the highlighted yellow area between two balls from the visual hint. An example of a text hint for the other task set could be seen in Figure 2-2. We invited the instructor (one of the co-authors) of the course that we recruited participants from to record audio hints since participants should be

familiar with his voice. The length of audio hints for all of the tasks was between seven to eight seconds long. This was the same duration as the visual hints and the text hint.

The multimedia hints were designed in accordance with most of Mayer's Multimedia Learning Principles (Mayer, 2005b) as follows:

Spatial and *Temporal Contiguity Principles*: If more than one hint were being provided for one task, such as in the visual + text, visual + audio, text + audio, or visual + text + audio conditions, different hints were presented spatially close to each other and displayed simultaneously. We put the text hint right above or below the place of the visual hint. Moreover, multiple hints were played at the same time.

Coherence Principle: All of the information presented in each of the conditions was relevant to the task. No extraneous text or information was provided to the students. When the multimedia hint was provided, the question text at the top of the slide was removed to avoid visual clutter.

Segmenting Principle: The hints were not presented in segments at a user-controlled pace. This was not deemed important in the context of our study because the length of each multimedia hint was only eight seconds and it could be viewed as many times as desired by the student.

Personalization Principle: The textual and audio hints used language and terminology that was familiar to the students and had been used in their class.

Voice Principle: The voice used in the audio version was the familiar voice of the instructor of the class.

Image Principle: No image of the speaker was used in the audio hints. Based on this principle, we decided that having the instructor's image would not affect learning in any way.

Pre-Training Principle: All students had covered material in the class that was relevant to the conceptual task and thus had received pre-training.

In summary, we used Mayer's multimedia design principles to design all of our hints. Further, we designed all of our hints to minimize the difference in information content, between the different conditions, such that the main difference between the conditions would be due to the different hint modalities.

Testing Hypotheses

As an experimental design with eight conditions, there could be as many as 28 pairwise comparisons. But not all of them are meaningful. Here we propose nine pairwise comparisons to test Mayer's Multimedia Learning Principles (Mayer, 2005b) and Wickens' MRT (2002) in our study.

There are two pairs of competing hypotheses we would like to test.

Competing Hypotheses I

Hypothesis I-A: The visual + text condition will outperform the text condition.

vs.

Hypothesis I-B: The visual + text condition will underperform the text condition.

Hypothesis I-A is supported by the *multimedia principle* of Mayer's CTML. It states, "Learners learn better when pictorial and linguistic information are presented together, rather than when linguistic information is presented alone" (Mayer, 2005b) According to this statement, adding visual hints to text hints would improve learning.

Hypothesis I-B is supported by Wickens' MRT. Presenting visual hints and text together is, according to Wickens (2002), "intra-modal time-sharing" (pp. 164). This type of time-sharing

would hinder participants' information processing. So, the text condition is predicted to have better performance than visual + text condition.

Competing Hypotheses II

Hypothesis II-A: The visual + text + audio condition will outperform the text + audio condition.

vs.

Hypothesis II-B: The visual + text + audio condition will underperform the text + audio condition.

Hypothesis II-A is supported by the *multimedia principle* because visual hints added to text and audio provide pictorial information along with linguistic information, while audio and text hints are linguistic information alone. Therefore, adding visual hints would improve participants' problem solving performance.

Hypothesis II-B, on the other hand is supported by Wickens' MRT. Adding visual cues interferes with other information presented in the visual modality, such as the picture and question of the problem, as well as the text cue. Therefore, the visual + text + audio condition will perform worse than the text + audio condition.

In addition to these two pairs of competing hypotheses, other hypotheses generated according to the principles of CTML and MRT are described below.

Hypothesis III: The visual condition will outperform the no hint condition.

Based on the *signaling principle*, cues that highlight the organization of the material improve learning. So participants who receive visual hints will outperform those that receive no hints.

Hypothesis IV: The audio condition will outperform the text condition.

This hypothesis is based on Mayer's *modality principle* and Wickens' MRT. Due to the *dual-channel* and *limited capacity assumptions* from CTML, and the fact that text and visual cues share the same modality in MRT, whereas visual and audio cues do not, narration facilitates learning better than on-screen text. Therefore, participants who receive audio hints will outperform those that receive text hints.

Hypothesis V: The visual + audio condition will outperform either the visual or the audio condition.

Both Mayer's CTML and Wickens' MRT model are in favor of this hypothesis. These two theories all prefer sending information via multiple channels rather than just one channel. The condition with visual + audio hints presents information by using both the visual channel and the auditory channel. So the participants from this condition would be expected to have better performance than either the visual or audio conditions.

Hypothesis VI: Either the text or the audio conditions will outperform the text + audio condition.

This hypothesis is generated by Wickens' MRT. Both text hints and audio hints use linguistic resources, so they will cause mutual interference when processed simultaneously. Therefore, the text + audio condition will underperform either text hints or audio hints.

Hypothesis VII: The visual condition will outperform the visual + text condition.

Wickens' MRT supports this hypothesis. Visual hints and text hints are both presented via the visual modality. Processing these two pieces of information may cause a heavy workload so that neither of them could be processed properly. Therefore, the visual + text condition would performance worse than the visual condition.

Hypothesis VIII: The visual + audio condition will outperform the visual + text condition.

26

According to Wickens' MRT, visual hints and audio hints use different modalities while visual hints and text hints use the same visual modality. Therefore, participants would benefit more learning from multiple modalities than from single modality.

Hypothesis IX: The visual + text OR audio conditions will outperform the visual + text + audio condition.

These comparisons test Mayer's *redundancy principle*. The core idea of the *redundancy principle* is "less is better than more". According to the *redundancy principle*, we hypothesize that the visual + text hint and visual + audio hint outperform the visual + test + audio hint, because they do not present redundant information to the learner.

A summary of the tested principles and hypothesized results is shown in Table 2-2.

Table 2-2 Nine pairs of comparison on testing Mayer's CTML and Wickens' model and the predicted results

	Pairwise Comparison	**Theoretical Basis**	**Hypothesized Result**
I	Visual + Text vs. Text	Multimedia Principle (CTML)	Visual + Text > Text
		Interference due to overloading of visual modality (CTML, MRT)	Visual + Text < Text
II	Visual + Text + Audio vs. Text + Audio	Multimedia Principle (CTML)	Visual + Text + Audio > Text + Audio
		Interference due to overloading of visual modality (CTML, MRT)	Visual + Text + Audio < Text + Audio
III	Visual vs. No hint	Signaling Principle (CTML)	Visual > No hint
IV	Text vs. Audio	Modality Principle (CTML, MRT)	Audio > Text

	Pairwise Comparison	Theoretical Basis	Hypothesized Result
V	Visual + Audio vs. Visual/Audio	Dual-channel assumption; no interference between visual and audio modality (CTML, MRT)	Visual + Audio > Visual/Audio
VI	Text/Audio vs. Text + Audio	Interference due to overloading of linguistic resources (MRT)	Text/Audio > Text + Audio
VII	Visual vs. Visual + Text	Interference due to overloading of visual modality (MRT)	Visual < Visual + Text
VIII	Visual + Audio vs. Visual + Text	Interference due to overloading of visual modality (MRT)	Visual + Audio > Visual + Text
IX	Visual + Text/Audio vs. Visual + Text + Audio	Redundancy Principle (CTML)	Visual + Text/Audio > Visual + Text + Audio

Experiment Procedure

Each participant in this study completed an individual session lasting about 45 minutes on average. A short oral explanation of the interview was given to each participant before the interview started. The explanation included the goal of this study, the procedure of the interview, a request for informed consent, and information regarding extra credit the participant would receive for their participation in the study.

Participants were randomly assigned to one of eight conditions: no hint (N = 20), visual only (N = 20), text only (N = 22), audio only (N = 21), visual + text (N = 18), visual + audio (N = 19), text + audio (N = 20), and visual + text + audio (N = 22). All participants solved all four task sets. Each task was presented on a computer screen. Participants were instructed to read the task carefully, view the hint when it was available, and then verbally provide their answer and

28

their reasoning supporting the answer to the interviewer when they were ready. In all of the seven hint conditions, there was a 10-second waiting time between the moments when the task was presented, and when participants could view the hint. We set this waiting time to avoid the participant rushing through tasks and hints without carefully reading the task. Participants had also been instructed that they could view hints as many times as they wanted. After participants provided their answer and reasoning to the interviewer, in some cases the interviewer asked some follow-up questions to clarify participants' answers or reasons. The interviewer took notes on participants' answers and reasons during the whole procedure of the interview. The entire interview session was audio and video recorded.

Results and Analysis

Scoring Procedure

The correctness of participants' responses was determined after all interviews were finished. Four raters completed the rating. Each of them was assigned to one task set to maximize consistency. To be coded as correct, a participant's response needed to have both the correct answer and correct reason. An incorrect answer or a correct answer with wrong reason would be graded as incorrect. Each grader graded 10 participants' interview notes with the help of videotapes. Afterward they discussed their ratings with the first author to have an agreement on a grading rubric for each task set. Then they graded all participants' responses for one task set separately. They marked the ambiguity responses. After they finished all grading assignments, the first author re-watched the videotapes for the ones that had been marked as difficult to rate, and re-rated them if necessary. The inter-rate reliabilities for each task set are listed in Table 2-3.

Table 2-3 Inter-rater Reliabilities for all task sets in all eight conditions

	Ball	Graph	Roller Coaster	Skier
Inter-rater Reliability	95.3%	98.3%	95.4%	96.6%
Missing Data	2.6%	1.9%	2.3%	2.1%

On some occasions, participants who were assigned conditions with hints accidentally gave the answers and reasons before they were presented with the hints and the interviewer did not remind the participant to view or listen to the hints. All of these responses were excluded from our data analysis, resulting in a small amount of data missing for each task set (see Table 2-3).

Initial and Transfer Task Performances

Initially, the performances on the initial tasks of each condition were analyzed. A Chi-square test was chosen to test the existence of a relationship between initial task performance and conditions. Four initial tasks from four task sets in our study functioned as a pretest. We would like to see how participants in each condition solved the initial tasks. The results of the Chi-square test showed no relationship between initial task performance and conditions, $\chi(21) = 22.484$, $p = 0.372$. This result verified that there was no statistically significant difference in performances on initial tasks between the conditions (see Table 2-4).

Table 2-4 Average correct solving percentages with std. error on initial tasks, near transfer tasks, and far transfer tasks over all four task-sets in all eight conditions

Condition	Initial	Near	Far
No Hint	10.0% $\pm$ 3.6%	18.8% $\pm$ 5.6%	25.0% $\pm$ 5.2%
Visual Only	11.3% $\pm$ 3.6%	32.5% $\pm$ 5.6%	31.3% $\pm$ 5.2%
Text Only	10.2% $\pm$ 3.4%	33.9% $\pm$ 5.3%	39.8% $\pm$ 5.0%

Condition	Initial	Near	Far
Audio Only	10.7% ± 3.5%	26.2% ± 5.5%	35.7% ± 5.1%
Visual + Text	4.2% ± 3.8%	30.6% ± 5.9%	33.3% ± 5.5%
Visual + Audio	15.8% ± 3.7%	34.2% ± 5.8%	31.6% ± 5.3%
Text + Audio	13.8% ± 3.6%	27.5% ± 5.6%	26.3% ± 5.2%
Visual + Text + Audio	6.8% ± 3.4%	30.7% ± 5.3%	34.1% ± 5.0%

Table 2-5 Average correct solving percentages with std. error on each training tasks over all four task-sets in all eight conditions

Condition	Training Problem						Average % Correct
	#1	**#2**	**#3**	**#4**	**#5**	**#6**	
No Hint	6.7% ± 6.3%	8.3% ± 6.5%	11.7% ± 6.6%	8.3% ± 6.8%	13.3% ± 6.8%	15.0% ± 6.9%	12.8% ± 4.9%
Visual Only	36.7% ± 6.3%	51.7% ± 6.5%	45.0% ± 6.6%	41.7% ± 6.8%	33.1% ± 6.8%	43.8% ± 6.9%	34.8% ± 4.9%
Text Only	28.7% ± 5.5%	37.5% ± 5.6%	32.0% ± 5.7%	37.6% ± 5.9%	36.1% ± 5.9%	39.3% ± 6.0%	33.0% ± 4.2%
Audio Only	23.6% ± 5.8%	19.4% ± 5.9%	27.8% ± 6.0%	23.1% ± 6.2%	27.0% ± 6.2%	30.9% ± 6.3%	24.9% ± 4.5%
Visual + Text	46.9% ± 6.1%	57.4% ± 6.3%	49.6% ± 6.4%	47.2% ± 6.6%	53.5% ± 6.6%	58.6% ± 6.7%	43.0% ± 4.7%
Visual + Audio	42.6% ± 6.0%	43.5% ± 6.1%	48.5% ± 6.2%	52.9% ± 6.4%	44.1% ± 6.4%	57.4% ± 6.5%	41.3% ± 4.6%
Text + Audio	37.5% ± 6.1%	29.2% ± 6.3%	28.4% ± 6.4%	31.1% ± 6.6%	32.8% ± 6.6%	39.2% ± 6.7%	30.0% ± 4.7%
Visual + Text + Audio	44.7% ± 5.5%	48.2% ± 5.6%	41.1% ± 5.7%	46.5% ± 5.9%	53.1% ± 5.9%	50.2% ± 6.0%	39.7% ± 4.2%

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that participants' performances on initial tasks, near transfer tasks, and far transfer tasks were significantly different, $F(2, 322) = 78.081$, $p < .001$. Post hoc tests using the Bonferroni correction revealed that there was a significant improvement from initial task performance to near transfer performance ($10.3\% \pm 1.3\%$ vs. $29.2\% \pm 2.0\%$, respectively, $p < .001$) and there was a significant improvement from initial task performance to far transfer performance ($10.3\% \pm 1.3\%$ vs. $32.3\% \pm 1.8\%$, respectively, $p < .0001$). However, there was a slight increase from near transfer task performance to far transfer task performance ($29.2\% \pm 2.0\%$ vs. $32.3\% \pm 1.8\%$, respectively), which was not statistically significant ($p = 0.356$). Therefore, we can conclude that the training process improves participants' performance in going from the initial to the transfer task.

As shown in Table 2-4, the initial task solving performance in the eight conditions are not exactly the same. To excluding the possible effect of initial task solving performance on any training tasks, near transfer tasks, or far transfer tasks, ANCOVA has been conducted in the following analyses with initial task performance as the covariant.

One-way ANCOVA with controlling initial task performance was completed to probe participants' performances on near transfer tasks and far transfer tasks in different conditions. The results indicated that there was no significant difference across all eight conditions on near transfer task performance, $F(7, 153) = 0.989$, $p = 0.442$; and far transfer task performance, $F(7, 153) = 1.449$, $p = 0.190$. The results seem to tell us that different hint modalities on the training tasks did not affect participants' performance on either the near transfer task or far transfer task significantly.

Three one-way ANOVA were conducted to separately compare the no hint condition's performance on initial tasks, near transfer tasks, and far transfer tasks with all hint conditions' performance on initial tasks, near transfer tasks, and far transfer tasks. We found there was no significant difference between no hint condition and hint conditions on initial task performance, $F(1, 646) = 0.11$, $p = 0.915$. There was a significant difference between no hint condition and hint conditions on near transfer task performance, $F(1, 646) = 4.814$, $p = 0.029$. There was no significant difference between no hint condition and hint conditions on far transfer task performance, $F(1, 646) = 2.198$, $p = 0.139$ (see Figure 2-3). Overall, these results showed hints could train participants solve transfer tasks better than no hint.

Figure 2-3 Initial task performance, near transfer task performance, and far transfer task performances of no hint condition and hit conditions. Error bars represent ±1 std. error of the mean.
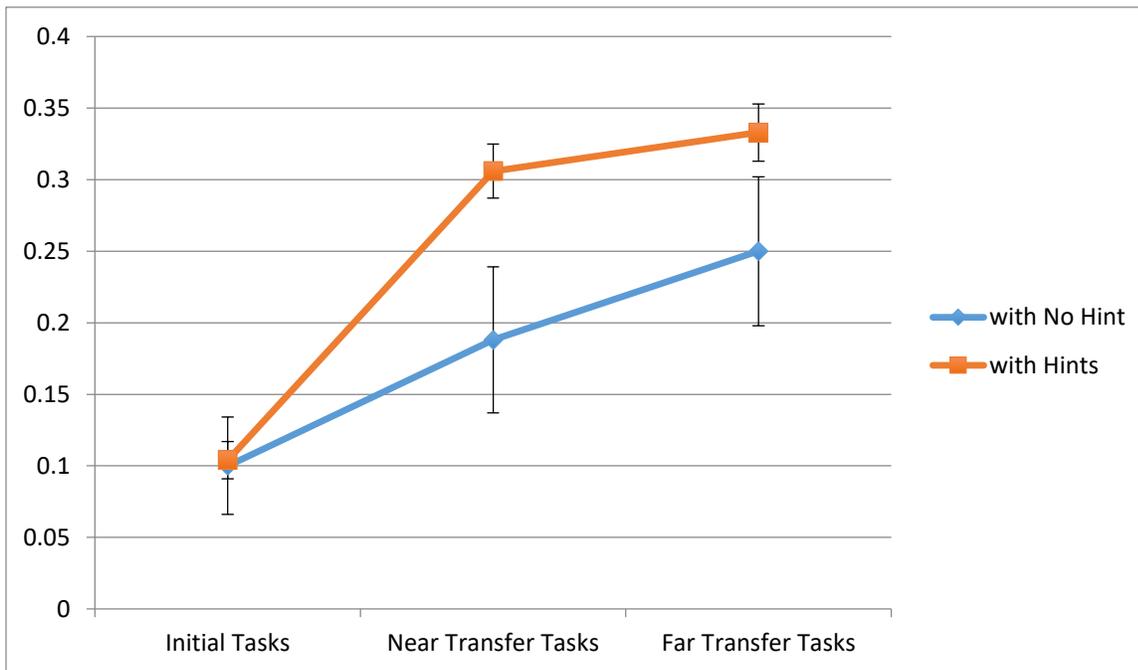
Training Task Performances

To understand how the participants improved from initial tasks to transfer tasks, their performances on the training tasks need to be analyzed. A one-way ANCOVA was used to

examine the training task performance between conditions. Figure 2-4 shows the means of correct solving percentage on training tasks for each of the design conditions. There was a significant difference between conditions ($F(7, 153) = 9.718$, $p < .001$). Therefore, we needed to examine the means of each condition to address our previously listed theoretical principles. Chi-square analysis has been used to probe those pairwise comparisons.

Figure 2-4 Average participant performance averaged across all task sets. Error bars represent ±1 std. error of the mean.
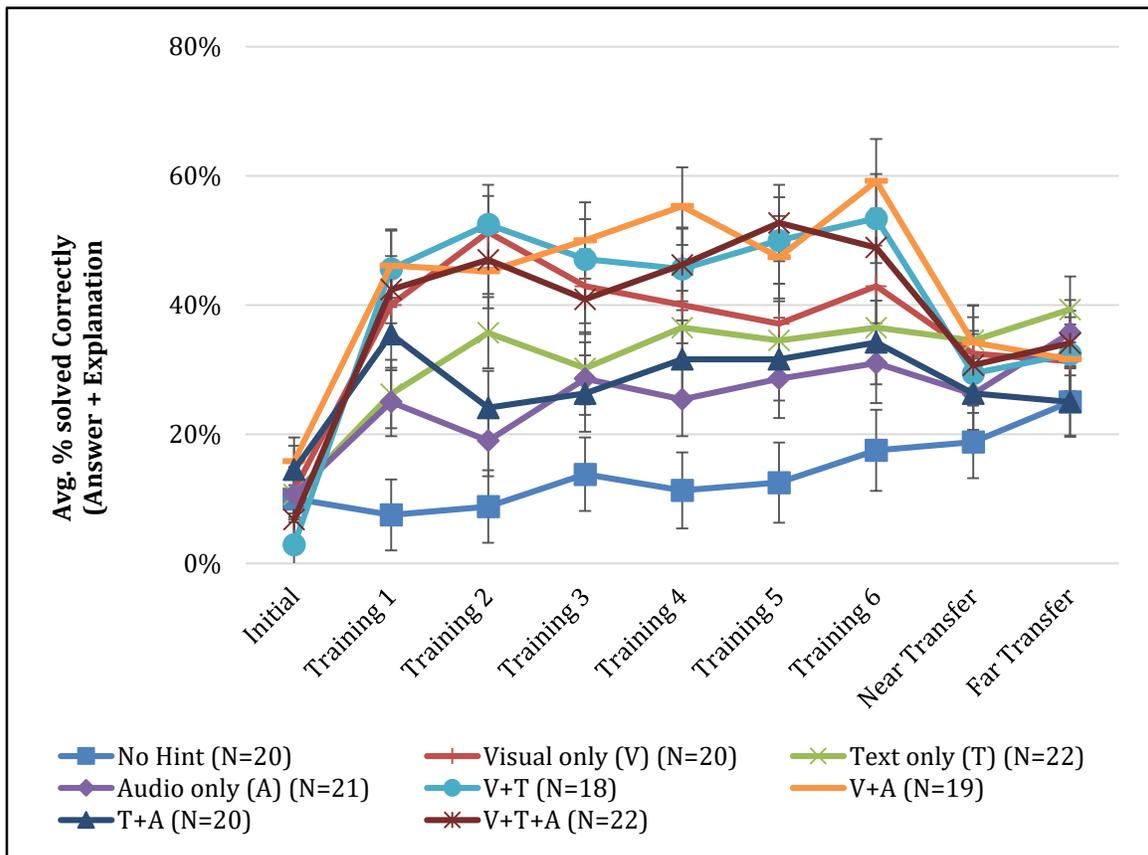
Pairwise Comparisons for Hypotheses

Hypotheses I-A and *I-B* compared the performance between the visual condition and visual +text conditions. There was a significant difference between the percentage of correct answers with visual + text hints and text hints ($\chi(1) = 22.092$, $p < .001$, 49.0% vs. 33.6%,

respectively). This was consistent with the *hypothesis I-A*, which was generated by the

multimedia principle of Mayer's CTML.

     Hypothesis II-A and *II-B* compared the performance between the visual + text + audio

and text + audio conditions. There was a significant difference between visual + text + audio

condition and text + audio condition ($\chi(1) = 24.763$, $p < .001$, 46.2% vs. 30.5%, respectively).

This result supported the *Hypothesis II-A*, which was generated by the *multimedia principle* of

Mayer's CTML.

     Hypothesis III was supported by the *signaling principle*. The visual hint condition

outperformed the no hint conditions ($\chi(1) = 113.134$, $p < .001$, 42.5% vs. 11.9%, respectively),

which supported the hypothesis.

     Hypothesis IV was supported by the *modality principle* that predicts that audio hints

would outperform text hints. Chi-square test showed that participants receiving text hints did

significantly outperform those receiving audio hints ($\chi(1) = 6.171$, $p = 0.013$, 33.6% vs. 26.4%,

respectively). This was contrary to the hypothesis.

     Hypothesis V was supported by the *dual channel assumption* of Mayer's CTML as well

as Wickens' MTR. It compared the visual + audio condition and visual or audio conditions.

There were two pairwise comparisons. There was a significant difference between visual

condition and visual + audio condition ($\chi(1) = 6.210$, $p = 0.015$, 42.5% vs. 50.7%, respectively).

And there was a significant difference between audio condition and visual + audio condition (χ

$(1) = 59.478$, $p < .001$, 26.4% vs. 50.7%, respectively). Both results supported the hypotheses.

     Hypothesis VI was supported by Wickens' notion of depressing performance due to

overloading of the linguistic resources. It predicted that either the text or audio conditions would

outperform the text + audio condition. Chi-square test showed there was no significant difference

between text condition and text + audio condition (χ (1) = 1.054, p = 0.305, 33.6% vs. 30.5%,).

Similarly, there was no significant difference between audio condition and text + audio condition (χ (1) = 1.981, p = 0.159, 26.4% vs. 30.5%, respectively). The text hints were marginally better than text + audio hints, which were marginally better than audio hints. These results did not support *Hypothesis VI.*

Hypothesis VII was supported by the notion that overloading the visual modality can depress performance. It predicted that the visual condition outperformed the visual + text condition, because the latter provided both visual and textual information using the visual modality, thereby overloading it. The result from the Chi-square test showed there was no significant difference between visual condition and visual + text condition (χ (1) = 3.745, p = 0.053, 42.5% vs. 49.0%, respectively). The visual + text condition outperformed the visual condition. These results did not support *Hypothesis VII.*

Hypothesis VIII was supported by the notion that overloading the visual modality can depress performance. It predicted that visual + audio condition would outperform the visual + text condition because the latter provides both visual and textual information using the visual modality, thereby overloading it. The Chi-square test showed that there was no significant difference between visual + audio condition and visual + text condition (χ (1) = 0.232, p = 0.630, 50.7% vs. 49.0%, respectively). These results did not support *Hypothesis VIII.*

Hypothesis IX had two pairwise comparisons based on the *redundancy principle*. Based on this principle, the visual + text + audio condition would outperform either the visual + text condition or the visual + audio condition. The Chi-square tests showed that there was no significant difference between visual + text + audio and visual + text (χ (1) = 0.690, p = 0.406, 46.2% vs. 49.0%, respectively) and similarly there was no significant difference between visual

+ text + audio and visual + audio (χ (1) = 1.813, p = 0.173, 46.2% vs. 50.7%, respectively). The comparisons showed the conditions with only two hint modalities did slightly better than the condition with all three hint modalities. These results did not support *Hypothesis IX*.

A summary of the results in light of the four principles is shown in Table 2-6. Bonferroni corrections were conducted for all the comparisons listed above. We found the pairwise comparisons for *Hypotheses I, II, III* still with the significant effects. While the pairwise comparisons for *Hypotheses IV* and *V* showed no significant differences with the Bonferroni corrections.

Table 2-6 Nine hypotheses and corresponding pairwise comparisons with statistical results

	Pairwise Comparison (conditions with average correctness)	p	Consistent with hypothesis?
I	Visual + Text (49.0%) vs. Text (33.6%)	< .001	*Hypothesis I-A*
II	Visual + Text + Audio (46.2%) vs. Text + Audio (30.5%)	< .001	*Hypothesis II-A*
III	Visual (42.5%) vs. No Hint (11.9%)	< .001	Yes
IV	Text (33.6%) vs. Audio (26.4%)	0.013	No[*]
V	Visual + Audio (50.7%) vs. Visual (42.5%)	0.015	Yes[*]
	Visual + Audio (50.7%) vs. Audio (26.4%)	< .001	Yes
VI	Text + Audio (30.5%) vs. Text (33.6%)	0.305	No significant difference
	Text + Audio (30.5%) vs. Audio (26.4%)	0.159	No significant difference
VII	Visual + Text (49.0%) vs. Visual (42.5%)	0.053	No significant difference
VIII	Visual + Audio (50.7%) vs. Visual + Text (49.0%)	0.630	No significant difference

	Pairwise Comparison (conditions with average correctness)	*p*	Consistent with hypothesis?
IX	Visual + Text + Audio (46.2%) vs. Visual + Text (49.0%)	0.406	No significant difference
	Visual + Text + Audio (46.2%) vs. Visual + Audio (50.7%)	0.173	No significant difference

[*] **The difference is not significant with Bonferroni correction.**

Discussion

The main purpose of this study was to explore the effects of different hint modalities and their combinations on facilitating learners to solve introductory level conceptual physics problems. Mayer's CTML (2002) and Wickens' MRT (2002) served as this study's main theoretical base and design guidelines. Since the theories have never been tested in the context of helping learners solve conceptual physics problems, the results of this study may suggest whether the applicable scope of CTML and MRT can accommodate this domain. We discuss three main outcomes of our results.

Reverse Modality Effect

Our study showed that audio hints were marginally less effective than text hints to help learners solve physics problems. Although the *reverse modality effect* has been discovered in several previous studies (Tabbers, *et al*., 2004; Leahy and Sweller, 2011), audio hints are still considered to be more efficient than text hints (Kalyuga, 2012) when the amount of time given to absorb information is limited. In other words, according to those previous modality studies, one might reasonably expect that audio hints are effective in facilitating physics problem solving. However, our results did not indicate any advantage for audio hints, compared to text hints. Not only that, there was evidence to suggest reverse modality effect in our data comparison, such that

the learners who received text hints outperformed the learners who received audio hints. The comparison we did was between the text only condition and the audio only condition. We also tested which hint modality, text or audio, would significantly improve visual hints. Therefore, we compared visual + text condition and visual + audio condition against visual only condition. Since we wanted to see what kind of hint would be more efficient for those who did not know how to solve the problems initially, we only included in these comparisons those participants who solved the initial problem incorrectly. We found that there was no significant difference between visual only condition and visual + audio condition (χ (1) = 3.671, p = 0.055, 38.1% vs. 44.8%, respectively). However, there was a significant difference between visual only condition and visual + text condition (χ (1) = 7.080, p = 0.008, 38.1% vs. 47.3%, respectively). These results suggested that text hints could improve the effect of visual hints better than audio hints. Many previous studies (Tabbers, *et al.,* 2004; Ginns, 2005; Stiller, *et al.,* 2009) found that the pacing mode of instruction is an important factor in modality effect. In our study, participants had been told they could view the hint as many times as they wanted to, but in only 3.13% of the cases, did the participants play the hint more than once. This means that an overwhelming majority of participants chose to stay in system-paced mode and the reverse modality effect that we found could not be explained by pacing mode of the hint. We suggest that this effect might be explained by preference of reading pace. Learners may prefer their own reading pace when they try to absorb a piece of abstract information. To be time-wise compatible the visual hint, audio hint, and text hint had been designed to be brief (one sentence with about ten words for each, see Figure 2-2). Each hint, regardless of condition, played for eight seconds. When the text hint was presented, learners might have been able to go back and forth between the terms or words that they tried to process. They were, of course unable to do this for the audio hints. This

feature of the hints in the text and visual modes might have helped learners better understand the hint, thereby making them more likely to integrate the information from the hint into their own problem solving.

Questioning the Dual-Channel Assumption

The *dual-channel assumption*, one of the fundamental assumptions of Mayer's CTML, needs to be tested in the context of conceptual physics problem solving. There were at least two pieces of evidence from our study that showed there might be more crosstalk between the visual and auditory channels than suggested by the *dual channel assumption* in Mayer's CTML.

First, we could not find any statistically significant support of the *redundancy principle*. The comparison between visual + text + audio and visual + audio conditions showed no statistically significant difference. Similarly, the comparison between visual + text + audio and visual + text conditions showed no statistically significant difference. We could use Mayer's CTML to explain the result of the former comparison as the capacity of the visual channel was not all occupied by processing visual hints. It was capable of processing the visual hint and text hint at the same time. But this went against the result of the latter comparison, since adding audio hints would definitely improve the combination of the visual and text hints according to this explanation, which is not consistent with what we found on the comparison between visual + text + audio and visual + audio. Therefore, Mayer's CTML may not be sufficient to explain the mechanism of attending hints and integrating the information with the brain activities for problem solving.

Second, we found that presenting audio hints in addition to visual hints could not improve performance over visual hints provided alone. This finding too is inconsistent with the *dual channel assumption*, especially when put with the fact that adding text hints with visual

40

hints could improve the performance with visual hints only. According to *dual-channel assumption*, audio hints with visual hints are the most effective combination since participants could process the information through auditory and visual channels in parallel. But the data from our study suggests that visual + text is the best combination (see Figure 2-3).

Mayer's CTML and Wickens' MRT

Mayer's CTML mainly discusses the visual and audio aspects of multimedia instruction. This is the *modality* dimension in Wickens' MRT. In addition to modality, Wickens' MRT also considers the coding dimension of the resource, i.e. verbal or spatial resources.

In this study, we have generated hypotheses to test both Mayer's CTML and Wickens' MRT in the domain of conceptual physics problem solving. We found that data from our study support Mayer's CTML when CTML and MRT make conflicting predictions (i.e., *competing Hypotheses I* and *II*). The hypotheses purely testing Wickens' model (i.e., *Hypothesis VI, VII,* and *VIII*) all showed no significant effect. However, Wickens' MRT can provide a different perspective on the multimedia hints. Its structural dichotomy, i.e., linguistic resource vs. spatial resource, may shed light in the further discussion around crosstalk between visual and auditory channels in perceiving on-screen text and narrated information.

In Chapter 4, we will expand the theoretical discussion by revisiting Mayer's CTML (2002) in light of frameworks of problem solving (Ohlsson, 1992) and Wickens' MRT (2002).

Chapter 3 - The Effect of Multimedia Hints on Participants' Eye Movements

Introduction

Chapter 2 discussed participants' problem solving performance in a study of training students to solve conceptual physics tasks with multimedia hints. The overarching goal of the study was to provide an experimental foundation for computer-aided instruction in regard to conceptual physics problem solving. The emphasis in Chapter 2 was to present the performance data and quantitative comparisons among eight conditions with different hint modalities. Unlike the hypotheses based on Mayer's (2005b) CTML, we found that the audio hints showed no advantage compared to the text hints. The condition with text hints outperformed the condition with audio hints. Moreover, presenting text hints with visual hints together significantly improved the performance of the visual hint condition. Presenting audio hints to learners helped students solve tasks better only when there was no other hint.

In this chapter, we focus on eye movement data collected in this study. The eye-tracking research has its own unique contribution to the study of problem solving and multimedia learning. Grant and Spivey (2003) studied participants' visual attention on solving Duncker's radiation problem (Duncker and Lees, 1945). They found that the participants who solved the problem correctly looked at the diagram differently from the participants who solved the problem incorrectly. Moreover, they also found that training participants that look at the part of the diagram where correct solvers attended, could significantly improve their problem solving performance. Their study showed visual attention is not a byproduct of the problem-solving process. Guiding participants' eyes can assist in guiding their thoughts. There are many studies along this line (Thomas and Lleras, 2007; 2009; Madsen, *et al.*, 2012; Rouinfar, 2014). In our

group's previous study (Rouinfar, 2014), we found that guiding participants' visual attention toward the expert-like area of a physics task's diagram, could improve participants' physics problem solving performance. There is a strong correlation between participants' visual attention, and problem solving performance. Further, going through a training process with visual hints can help participants solve transfer tasks more efficiently than those being trained through a process without visual hints. Therefore, guiding visual attention during problem solving can also improve physics problem solving performance (Rouinfar, *et al.,* 2014).

On the other hand, visual attention can grant us insight in studying multimedia learning. Mayer (2010) valued eye-tracking study as the opportunity to understand the perceptual process during multimedia learning to explain how multimedia instruction works. Before merging of eye-tracking methodology and multimedia learning study, multimedia instruction could only be evaluated by measurements such as pre- and post-test, retention test, and self-report survey. None of them could provide real-time evidence to reveal what is happening when a learner is viewing the multimedia instruction. Now with the eye-tracking technology, the time period during multimedia instruction is less of a black box to researchers.

Due to these empirical benefits of visual attention studies, we focus on participants' eye movement data in this paper and try to gain insights concerning the problem solving performance data presented in Chapter 2. Chapter 2 presented a discussion on what works. In this paper, we focus on why it works with the evidences from the analysis of eye movement data.

Theoretical Background

Eye-tracking Studies on Multimedia Learning

Eye-tracking technology has been recognized as a unique tool for studying multimedia learning. In this group of studies, researchers usually were interested in two questions: (1) What

areas did participants look at? (2) For how long did participants look at those areas? Schmidt-Weigand and his colleagues (2010a) carried out two eye-tracking experiments to studying the modality effect proposed by Mayer's CTML (Mayer, 2002). According to CTML, splitting attention is the main reason for the modality effect. Learners may miss the important textual information when they are viewing visualizations and vice versa. There are controversies around the modality effect. The reverse modality effect has been reported from a study of participants using self-paced instruction (Stiller, *et al*., 2009). Schmidt-Weigand's experiments were designed to probe the interactions between instructional paces, and information modalities (i.e., on-screen text and narration) on participants' visual attention. They found participants in their study spent longer time on reading the on-screen text than viewing visualizations. Even though they did not find a very clear modality effect (i.e., the participants with narration did not outperform the participants with on-screen text), their attempt to use eye-tracking technology benefited the study of multimedia learning. The *Signaling Principle* is another multimedia design principle that has been addressed in many eye-tracking studies. Both Boucheix and Lowe's study (2010) and de Koning *et al.*'s study (2010) found that highlighting the relevant features could draw participants' attention to those features and improve their learning outcomes. Ozcelik *et al.* also studied the signaling effect on multimedia learning with the support of eye movement data (2010). They found that highlighting the relevant features not only drew participants' attention to those areas, but also improved the efficiency of finding relevant information.

Those studies normally used "dwell time" as a measure of participants' visual attention. Dwell time is the amount of time that participants spend looking at a certain area (i.e., areas of interest, AOI). There are two potential issues. First, the total viewing time should be taken into account. For example, a one-second dwell time on the AOI#1 within a total viewing time of two

seconds should be different from a one-second dwell time on AOI#2 within a total viewing time of twenty seconds. Second, the pixel area of AOI should also be considered. The dwell time on a small AOI should be weighted more heavily than the dwell time on a large AOI since the chance that fixations fall onto a large AOI is higher than the chance that fixations fall onto a small AOI. To resolve these issues, a new measure, the percentage of dwell time divided by the percentage of pixel area was proposed (Fletcher-Watson *et al.*, 2008). This measure is called the *domain relative ratio*.

Our group's previous studies have also benefited from using eye-tracking technology. A problem-solving study that recorded eye movements (Rounifar, *et al.*, 2014) found that providing participants with visual hints over multiple task trials could help develop automaticity of extracting information from the diagrams' relevant areas later on when no visual hints were presented. A follow up study (Agra, 2015) used patterns emerging from eye movement data together with problem solving accuracy and confidence, and the provision of feedback, to suggest when problem-solving impasse occurred, and the mechanisms of breaking such impasses.

Cognitive Load Theory and Measurement of Eye Movements

Mayer's CTML is a theory built on cognitive studies in guiding multimedia instructional design (Mayer, 2005a; 2005b). In his work, Mayer defined the term "multimedia" as "presenting words (such as printed text or spoken text) and pictures (such as illustrations, photos, animation, or video)" (pp. 2). The theory is to model the mechanisms by which humans learn from such multimedia information.

Three theoretical assumptions serve as the keystone of Mayer's theory: the *dual-channel assumption*, the *limited capacity assumption*, and the *active-processing assumption.* Among

these three assumptions, the *dual-channel assumption* and the *limited capacity assumption* have helped generate many testable hypotheses and practical instructional design principles. The *dual-channel assumption* originated from the *sensory-modalities* approach (Mayer, 2002). *Cognitive load theory* provides a theoretical foundation for the *limited capacity assumption*. To fully understand CTML, *cognitive load theory* needs to be carefully explored. *Cognitive load theory* (Sweller, 1988; 1989) has stated that the instructional techniques are likely to fail if they require a processing capacity beyond learners' limits. The idea of people's limited processing capacity came from Miller's discussion on the "magic number seven" (Miller, 1956) and de Groot's study on chess players (de Groot, 1978). Miller included *absolute judgments* of auditory pitch, auditory loudness, taste intensity, and the pointer's position and summarized that to the human brain, the numbers of distinguishable categories on these stimuli are always around seven. It suggested that our brains had some limitation on the capabilities of making judgment. de Groot (1978) found there was a huge difference between chess masters and novices with regard to their ability of reconstructing a chess position. This was due to a master's superior domain knowledge, not a better visual short-term memory since masters and novices were equally poor when the pieces were placed randomly. In a more recent study, Cowan (2001) revisited the mental storage capacity of short-term memory and found that people can only hold about four chunks in short-term memory. All of these evidences suggest that when one was engaging in a task (e.g., problem solving, comprehension, or memorization), there is a "bottleneck" in his/her cognitive system. The effort needed to process information through this bottleneck could be qualitatively and quantitatively measured in terms of cognitive load, based on *cognitive load theory*.

As shown in earlier studies, human's processing capacity is limited. It is expected that the instructional material with less complexity will impose low cognitive load on humans'

information processing systems. Instructional material's complexity is due, in part, to its element interactivity. According to Ginns' work (2005), element interactivity is "the extent to which the learning task requires the student to hold several related chunks of to-be-learned information in working memory simultaneously in order to comprehend then learn the concept or procedure" (pp. 320). This type of cognitive load caused by characteristics of learning materials is called *intrinsic cognitive load* (Paas, *et al.,* 2003).

Intrinsic cognitive load (Paas, *et al.,* 2003) represents an instructional material's inherent complexity. Yet, instruction using the same material may impose different amounts of intrinsic cognitive load on different learners. For example, instruction about physical dynamics may include items involving concepts such as: kinetic energy, potential energy, time, speeds in different directions, acceleration in different directions, and moving distances/displacements in different directions. Each item is related to other items. However, to a novice physics learner, the *intrinsic cognitive load* of this instruction is high since he/she needs to keep track of multiple items in order to understand one item. However, the intrinsic cognitive load for an expert physics learner on this instruction could be low since he/she is familiar with these relationships. To study each item in the instructional material, the expert just needs to focus on the item and the relationships connecting this item to other items. Such cognitive structures, which connect items are called *schemas* and *schemas* could reduce *intrinsic cognitive load*.

Other than *intrinsic cognitive load*, there are another two categories of cognitive load, according to Paas, *et al.* (2003). The *germane* or *effective cognitive load* is the type of cognitive load that is imposed by activities directly related with the task at hand. This is the type of cognitive load that enhances learning. The last category of cognitive load is the *extraneous* or *ineffective cognitive load*. This is the type of cognitive load caused by activities that are not

helpful to learning. Unlike *intrinsic cognitive load*, these two types of cognitive load could be modified by instructional designers. The main purpose of many of Mayer's multimedia instruction design principles (Mayer, 2005b), such as the *spatial contiguity principle*, the *temporal-contiguity principle*, and the *coherence principle,* is to reduce the *extraneous cognitive load* by optimizing the multimedia instructional design.

The fact that the *Cognitive load theory* correlates learning outcomes and cognitive load shows the necessity of reliable measurements of cognitive load. Paas and van Merrienboer's model (1994) suggested that the measurements of mental load, mental effort, and task performance, could indicate cognitive load. Researchers often rely on self-report surveys and questionnaires to measure participants' mental burden. Paas (1992) used a mental-effort rating scale to measure a participants' cognitive load while studying statistical problems and achieved plausible results. After Paas's successful attempt, the rating scale has become a popular cognitive load measurement technique, which has been used in many studies (Kalyuga, *et al*., 1998; Mayer and Chandler, 2001; Tabbers, *et al*., 2004).

However, even though the above rating scale has been demonstrated as a valid, accurate, and sensitive measurement of cognitive load, it still relies on self-reported information from the participants. Some researchers prefer techniques that use physiological measurements with less subjective bias. Tracking eye movements is one of the techniques that have been used to objectively measure cognitive load.

Early eye movement studies revealed two major components of eye movements: *saccades* and *fixation pauses*. A saccade is the rapid oculomotor movement of eyes (Javal, 1879). Visual perception of the external input is significantly reduced during the time of saccade (Burr & Morrone, 1996; Burr, Morrone, & Ross, 1994, October 6; Chekaluk & Llewellyn, 1994;

48

Paus, Marrett, Worsley, & Evans, 1995; Volkmann, 1986). A fixation (pause) is between

saccades when eyes are relatively fixed at a given location to gather the information there. The

mean saccade length tends to vary from 1 degree to 4 degrees of visual angle and the mean

fixation duration generally varies from 200ms to 400ms (Rayner, 1998).

There were two competing hypotheses in understanding the relationship between human

eye movements and cognition. The first one was the *cognitive-lag hypothesis* (Kolers, 1976).

According to this hypothesis, the cognitive process is not quick enough to affect eye movements.

Eyes just serve as the entrance of the external visual information. All complex cognitive

activities happen beyond the movements of the eyes. A competing hypothesis was the *process-

monitoring hypothesis* (Rayner and McConkie, 1976). This hypothesis suggested that the

cognitive processes happen during the time frame of a fixation. In other words, the fixation

duration is a function of cognitive load. The latter hypothesis aligned with many theoretical and

experimental studies. Just and Carpenter (1980) developed the eye-mind model and applied it to

their theory of reading. The eye-mind model suggested that real-time cognitive processing was

immediate and low-cost. Studies on reading text (Rayner & McConkie, 1976), attending to

subtitles (d'Ydewalle, *et. al*., 1991), and viewing pictures (Underwood *et. al.,* 2004) all

supported the process-monitoring hypothesis and showed a close relationship between fixation

duration and the real-time cognitive processes (for review, see Rayner, 1998).

Some studies have found that the measure of fixation duration did not align with other

cognitive load measurement techniques. Van Gog *et al*. (2005) found the high expertise

participants had longer mean fixation durations in some phases of tasks comparing with the low

expertise participants. Amadieus *et al*. (2009) found that the subjective ratings of mental effort

were not always consistent with the measure of fixation durations. But in those studies, levels of

prior knowledge determined how students felt about the task. And ratings of mental effort reflected the overall task difficulties. Fixation durations measure cognitive load within a certain time frame. Van Gog and his colleagues suggested that these two cognitive load measurements measured different aspects of cognitive load was the reason of this mismatch (Van Gog, *et al*., 2009).

Multiple Resources Theory and Multitask Performance

According to *cognitive load theory* (Sweller, 1988), creating high cognitive load should negatively impact problem-solving performance. However, it does not provide a clear map that could directly relate the properties of external information to cognitive load, especially when there are multiple sources of external information provided to the learner, the information processing needs to be clarified. Wickens (2002) proposed a theoretical model, which is called *multiple resources theory* (MRT), to predict such time-sharing ability.

MRT was influenced by many early multitask studies (Bahrick, *et al*., 1954, Briggs, *et al*., 1972; North and Gopher, 1976). Compared to Mayer's CTML (2002), which mainly discusses the modalities of instruction, Wickens' MRT provides three more dimensions to fit the wider range of external information in real world's task performance, such as driving a car (Palinko, *et al*., 2010) and flying an aircraft (Wickens, *et al.,* 2003). The *resources* aspect of MRT is aligned with mental workload, which in other words, is cognitive load. MRT argues that the processing resources of the brain can be separated into four dimensions: perceptual *modalities* (visual and auditory), processing *stages* (perception and cognition, then response), visual *channels* (focal and ambient), and processing *codes* (linguistic and spatial). Task performance should not be adversely affected if the cognitive tasks are not overlapping in these dimensions. MRT provides a coherent explanation for many experimental results and is

consistent in some ways with Mayer's CTML. Besides the driving scenarios which have been

mentioned earlier, it can be used to explain the principles of multimedia instruction design

(Mayer, 2005b). For example, pictures contain spatial information conveyed by the visual

modality. Narrations contain linguistic information conveyed by the auditory modality. Text

contains linguistic information conveyed by the visual modality. To learners, pictures and

narrations together should be a better combination than pictures and texts, since pictures and

texts together would overload information from the visual modality. This is consistent with the

modality principle from Mayer's CTML (2002). Moreover, the interference between two pieces

of information that both use linguistic resources can explain the disruption of office noise

withspeech on memory for prose and task performance (Banbury and Berry, 1998).

On the other hand, some other studies have reported results which contradict Wickens'

MRT. Latorella (1998) found the effect of preemption on studying pilot performance in a flight

simulator. The auditory information was found to attract pilot's attention away from visual

information and weaken the performance, even though there should be no cognitive load

overload in any dimensions according to Wickens' MRT. Spence and Driver (2000) suggested

that auditory information may dominate visual information due to the transient nature of auditory

information, and therefore the need for the learner to attend to the auditory information rather

than the visual information.

Significance of Study

In summary, previous studies showed that eye-tracking technology can make a unique

contributions in studies of learning with multimedia materials. Eye movements can reveal the

way participants perceive hints that are provided to facilitate problem solving. Analyzing eye

movement data can help us explain the problem solving performance data reported in Chapter 2. There are two research questions guiding our eye movement study.

1. How does presenting hints in different modalities affect participants' attention during problem solving?

2. How does presenting hints in different modalities affect the cognitive load during problem solving?

Method

The Eye-Tracking Technology

The detailed descriptions of the participants, experimental procedure, and interview materials can be found in Chapter 2. Here we focus on describing the eye-tracking aspects. The participants' eye movements were recorded by using either the EyeLink 1000 or EyeLink 1000 Plus systems (http://www.sr-research.com). According to the requirements of our study, both systems had the necessary spatial and temporal resolution for our study.

The problems, and visual and/or text hints were presented on a computer screen with 1024×768-pixel resolution and 85 Hz refresh rate. We used chin and forehead rests to stabilize participants' heads and minimize the error of eye movement recording. The chin and forehead rests were set 24 inches away from the computer screen. If the eye's acceleration exceeded $8,500^{\circ}/s^2$ and speed exceeded $30^{\circ}/s$, the eye tracker identified this eye movement as a saccade. Otherwise, it counted as a fixation. We focused on the eye movement data on solving training tasks since the problem solving performance on training tasks showed significant differences between hint modality conditions

A valid eye movement analysis should define a certain time period and a certain area. The time period in solving each training task is divided into three time sections: before hints,

during hints, and after hints. On each training task with hints, hints were not allowed in the first ten seconds. This amount of time was given to participants to read the statements and inspect the graphs. Then they were allowed to press a designated button on the control box to call out the hints. The duration of each hint section was set to be eight seconds. Participants could view hints as many times as they wanted to. But in a vast majority (96.87%) of instances, participants played the hints only once. After viewing the hints, they were asked to verbally report their answers to interviewer. The after hints period was for 10 seconds, starting immediately after the hint disappeared from the screen.

Areas of interest (AOIs) were drawn with a border of $1.1°$ of visual angle (see Figure 3-1) around the visual hints and text hints associated with each training task. As discussed in the earlier section, domain relative ratio (Fletcher-Watson, *et. al*., 2008) was calculated by normalizing the percentage of viewing time spent in each AOI with the percentage of corresponding AOI's area relative to the pixel area of the whole computer screen.

Figure 3-1 An example of drawing visual hint AOI and text hint AOI in a slide with hints.
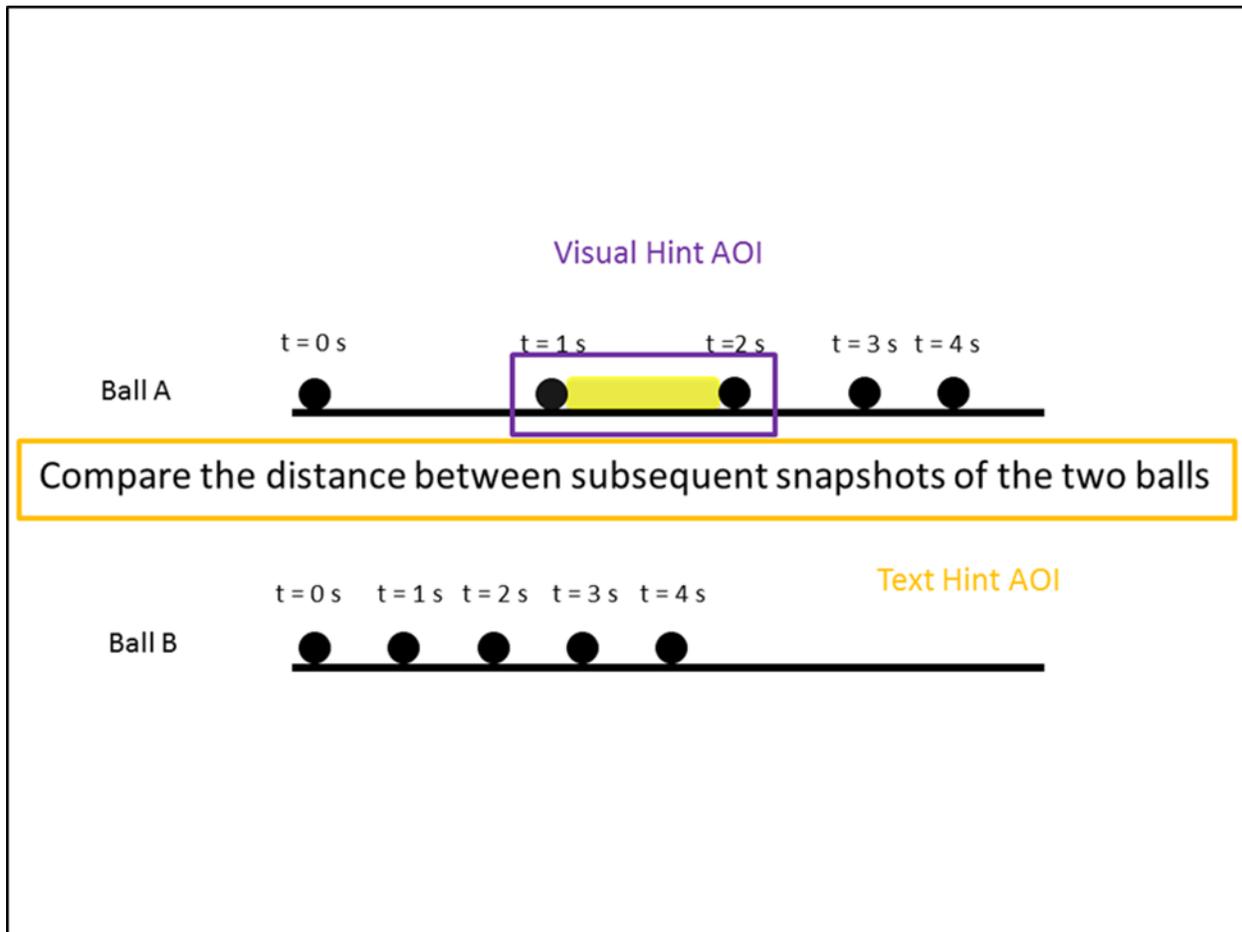
Testing Hypotheses

The domain relative ratios and mean fixation durations are two types of eye movement data that are analyzed in this study. As the guidance of data analysis, hypotheses are generated based on the theories and empirical evidences.

Hypotheses Regarding Domain Relative Ratios Comparisons between Conditions

There are four hypotheses for the domain relative ratios.

Hypothesis I: Adding visual hints with text hints would make participants less likely to view text hints and vice versa.

Many studies reported the *split-attention effect* when pictorial information and on-screen text information are presented simultaneously (Mayer and Moreno, 1998; Kalyuga, *et al.,* 2011). Participants have to choose between viewing visual hints and reading text hints. Therefore, the amount of time participants spent to view visual hints would decrease when text hints were on screen. It also would be true for reading time on text hints.

Hypothesis II: Participants would likely be biased to view text hints when text hints and visual hints were presented simultaneously.

Previous eye movement study on multimedia learning showed that learners preferred reading text to viewing pictorial information (Schmidt-Weigand, *et al*., 2010b). Once text information was presented, learners' attention would be attracted to read the content. According to this empirical evidence, we predict that the reading time on text hints would exceed the viewing time on visual hints when the text and visual hints were presented together.

There are also two pairs of competing hypotheses.

Competing Hypotheses III

Hypothesis III-A: Participants in the visual + audio condition would view visual hints for longer time than those from visual condition.

Hypothesis III-B: Participants in the visual + audio condition would view visual hints for shorter time than those from visual condition.

Hypothesis III-A is supported by the *preemption effect* (Latorella, 1998). According to this effect, participants would mainly focus on hearing audio hints when audio hints and visual hints were presented at the same time. Participants may need more time to perceive visual hints when there are no audio hints.

Hypothesis III-B is supported by Wickens' MRT (2002) predicts differently than above. According to MRT, we do not anticipate interference for processing audio hints and visual hints in parallel, because the audio hints use linguistic resources, while video hints use visual resources. Moreover, participants with visual + audio hints could obtain help information from hearing the audio part, they may inspect visual hints for shorter time than those with visual hints only.

Competing Hypotheses IV

Hypothesis IV-A: Participants in the text + audio condition would view text hints for longer time than those from text condition.

Hypothesis IV-B: Participants in the text + audio condition would view text hints for shorter time than those from text condition.

Hypothesis IV-A is supported by Wickens' (2002) MRT which would categorize text hints and audio hints as linguistic information which would be processed by linguistic resources. Thus interference is expected between text and audio hints. The participants with text + audio hints need to spend more time to read text hints than the participants with text hints.

Hypothesis IV-B is consistent with empirical research by d'Ydewalle *et al.* (1991) who studied how people watched a film with subtitles. They found their participants read subtitles for marginally more time when the sound was muted.

Hypotheses Regarding Mean Fixation Duration Comparisons between Conditions

Fixation duration is another important measure used in the analysis of eye movement data. For the analysis on mean fixation duration, we focus only on participants who solved the first training task correctly, after solving the initial tasks incorrectly. There are two reasons. First, the first training task on each task set was the first time participant was exposed to hints. The

cognitive load during hints on the first training task reflected how participants attended hints the best. Second, participants who solved initial tasks correctly knew how to solve tasks without viewing hints. The cognitive load from those who solved initial tasks incorrectly indicated the effort participants exerted to integrate hint information with problem solving.

Cognitive load theory relates the lower cognitive load to higher task performance (Sweller, 1988; 1989). The performance data presented in Chapter 2 showed the following differences in performance

- visual hint condition outperformed the no hint condition.

- visual + text hints condition outperformed the text hint condition;

- visual + audio hints outperformed the audio hint condition.

- visual + text + audio hints outperformed text + audio hints;

- text hints outperform audio hints; and

- visual + audio hints outperformed visual hints.

The analysis including the participants who solved initial problems incorrectly showed that the visual + text hints outperformed the visual hints. The hypotheses below mainly test the comparisons that were found to have significant differences with regard to problem solving performance.

Hypothesis V: Participants with text hints have shorter fixation durations than those with audio hints.

Hypothesis VI: Participants with text hints have shorter fixation durations than those with text + audio hints.

These two hypotheses are suggested by the performance data reported in the previous paper of this two-paper sequence. There was only a marginal difference between the text

condition and text + audio conditions. Another support for *Hypothesis VI* is provided by Wickens' (2002) MRT, which clearly predicts cognitive overload in processing text and audio information in parallel. Thus, these hypotheses are generated by both experimental evidence and theoretical prediction.

Next there are two pairs of competing hypotheses.

Competing Hypotheses VII

Hypothesis VII-A: Participants in the visual + text and visual + text + audio conditions will have shorter fixation durations than those in the text and text + audio conditions respectively.

Hypothesis VII-B: Participants in the visual + text and visual + text + audio conditions will have longer fixation durations than those in the text and text+ audio conditions.

It has been shown that conditions with visual hints outperformed conditions without visual hints. *Hypothesis VII-A* is supported by the notion that the better performance of the visual hint conditions was due to lower cognitive load as measured by mean fixation duration. However, there were studies showing the mean fixation durations in reading and scene perception are different (Rayner, 1998). The mean fixation duration in silent reading is about 225ms and the mean fixation duration scene perception is about 330ms. *Hypothesis VII-B* is supported by the notion that presenting text hints on-screen would change participants' behavior from scene perception to reading, which would result in shorter fixation durations with text hints and text + audio hints compared to visual + text hints and visual + text + audio hints.

Competing Hypotheses VIII

Hypothesis VIII-A: Participants in the visual hint condition have shorter fixation duration than those in the visual + audio hint conditions.

Hypothesis VIII-B: Participants in the visual hint condition have longer fixation durations than those in the visual + audio hint conditions.

Hypothesis VIII-A is supported by the preemption effect (Latorella, 1998). Adding audio hints to visual hints would draw participants' attention away from visual hints, which would result in processing visual hints slower than without audio hints. However, the performance data presented in Chapter 2; showed that the visual + audio condition outperformed the visual condition. *Hypothesis VIII-B* is supported by the notion that this improvement is due to reducing cognitive load. However, it must be noted that the participants who solved the initial problem incorrectly did not have significant difference in problem solving performance between visual + audio condition and visual condition.

Results and Analysis

Evidence of Split Attention

A two-way 4 (conditions with visual hints, i.e., visual condition, visual + audio condition, visual + text condition, and visual + text + audio condition) × 6 (six training tasks) mixed ANOVA with repeated measures on training tasks was conducted to compare domain relative ratio on visual hint AOI during hints. There was a significant main effect of training tasks, $F(5, 315) = 3.843$, $p = 0.002$. And there was a significant main effect of conditions, $F(3, 63) = 14.699$, $p < .001$. These two main effects were qualified by the significant interaction, $F(15, 315) = 2.255$, $p = 0.008$ (see Figure 3-2). To explore participants' visual attention change in the entire training process, domain relative ratios were compared across all six training tasks for each condition. Four one-way repeated measures ANOVA were conducted. We found a significant effect of training tasks in the visual condition, $F(5, 70) = 4.896$, $p = 0.004$; and visual + audio condition, $F(5, 80) = 2.639$, $p = 0.050$. There is no significant effect of training tasks in the

visual + text condition, $F(5, 70) = 1.915$, $p = 0.128$; and visual + text + audio condition, $F(5, 95) = 0.670$, $p = 0.597$.

Figure 3-2 The domain relative ratios on the visual hint AOI on the training tasks. Error bars indicate ±1 standard error of the mean.
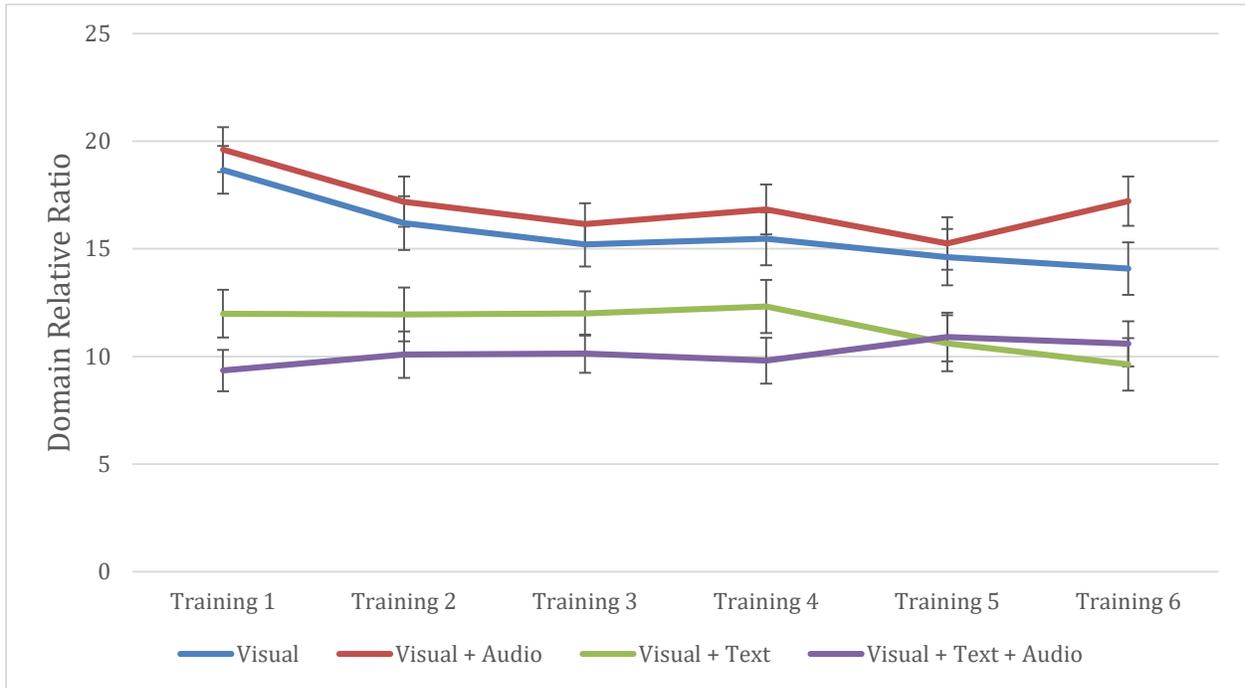

Figure 3-3 The training task solving performances of visual, visual + audio, visual + text, and visual + text + audio conditions. Error bars indicate ±1 standard error of the mean.
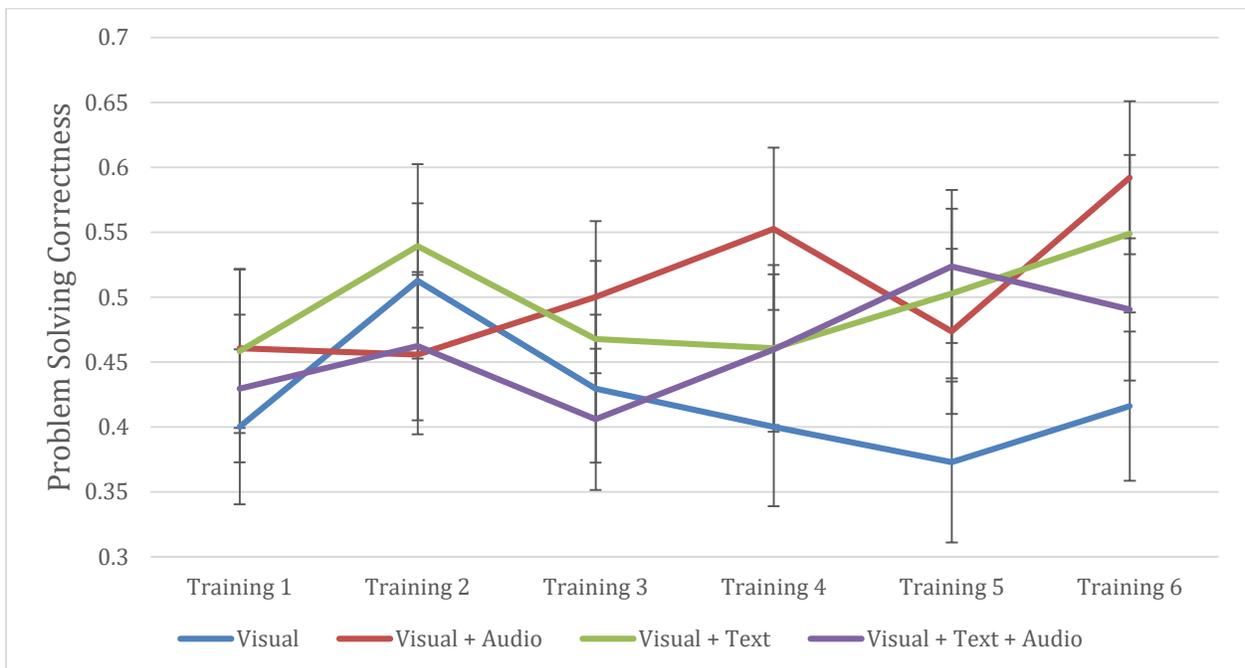
To further probe the interaction, in order to compare the way that hint modalities affect participants' visual attention on each training task, domain relative ratios were compared across four conditions in each training task. Six one-way ANOVA were conducted. All of them showed significant effects of conditions (see Table 3-1). Post hoc pairwise comparisons with Bonferroni correcting found that significant differences occurred only between the conditions with text hints (i.e., visual + text and visual + text + audio) and the conditions without text hints (i.e., visual and visual + audio). There is no significant difference for adding audio hints or not (see Table 3-2).

Table 3-1 The *F* values and *p* values of comparing the domain relative ratio on the visual hint AOI between visual condition, visual + text condition, visual + audio condition, and visual + text + audio condition for each training task.

Training Task	#1	#2	#3	#4	#5	#6
F	23.959	8.652	8.790	7.904	3.859	9.042
p	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$

Table 3-2 The *p* values from pairwise comparisons with Bonferroni correction on the domain relative ratio on visual hint AOI between conditions for each training task

Training Task	#1	#2	#3	#4	#5	#6
Visual vs. V + T	$< .001$	0.115	0.188	0.456	0.203	0.073
Visual vs. V + A	1.000	1.000	1.000	1.000	1.000	0.391
Visual vs. V + T + A	$< .001$	0.003	0.003	0.006	0.211	0.205
V + T vs. V + A	$< .001$	0.020	0.027	0.058	0.071	$< .001$
V + T vs. V + T + A	0.462	1.000	1.000	0.764	1.000	1.000

V + A vs. **V + T + A**	< .001	< .001	< .001	< .0001	0.067	< .0001

Another two-way 4 (conditions with text hints, *i.e.,* text condition, text + audio condition, visual + text condition, and visual + text + audio condition) × 6 (six training tasks) mixed ANOVA with repeated measures on training task was conducted to compare domain relative ratio on text hint AOI during hints (see Figure 3-4). There was a significant main effect of training task, $F(5, 340) = 34.866$, $p < .001$, and a significant main effect of condition, $F(3, 68) = 24.732$, $p < .001$. There was no significant interaction, $F(15, 315) = 1.491$, $p = 0.116$ (see Figure 3-4). Post hoc texts with the Bonferroni correction were conducted to probe these two main effects. The results were summed up in Table 3-3 and Table 3-4. We found that presenting visual hints with text hints would significantly decrease the viewing time on text hints. Adding audio hints does not have such an effect. Moreover, the viewing time on text hints was decreased from the first training tasks to the last training tasks.

Figure 3-4 The domain relative ratios on the text hint AOI on the training tasks. Error bars indicate ±1 standard error of the mean.
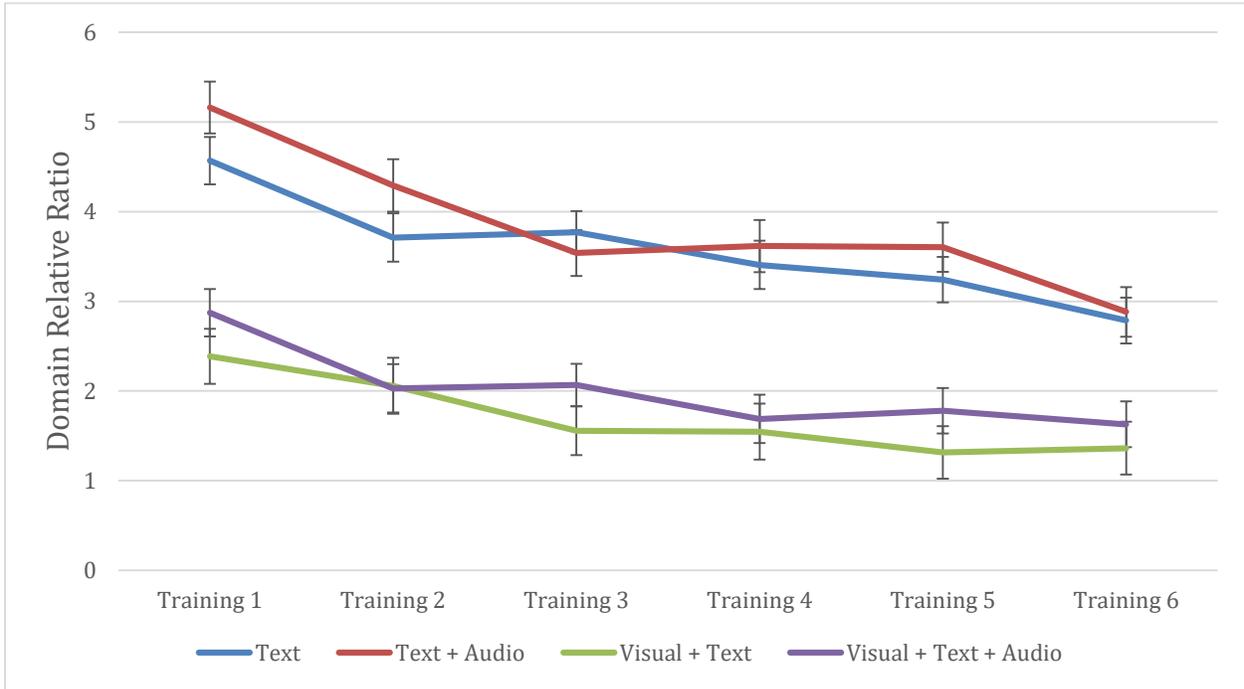

Figure 3-5 The training task solving performances of text, text + audio visual + text, and visual + text + audio conditions. Error bars indicate ±1 standard error of the mean.
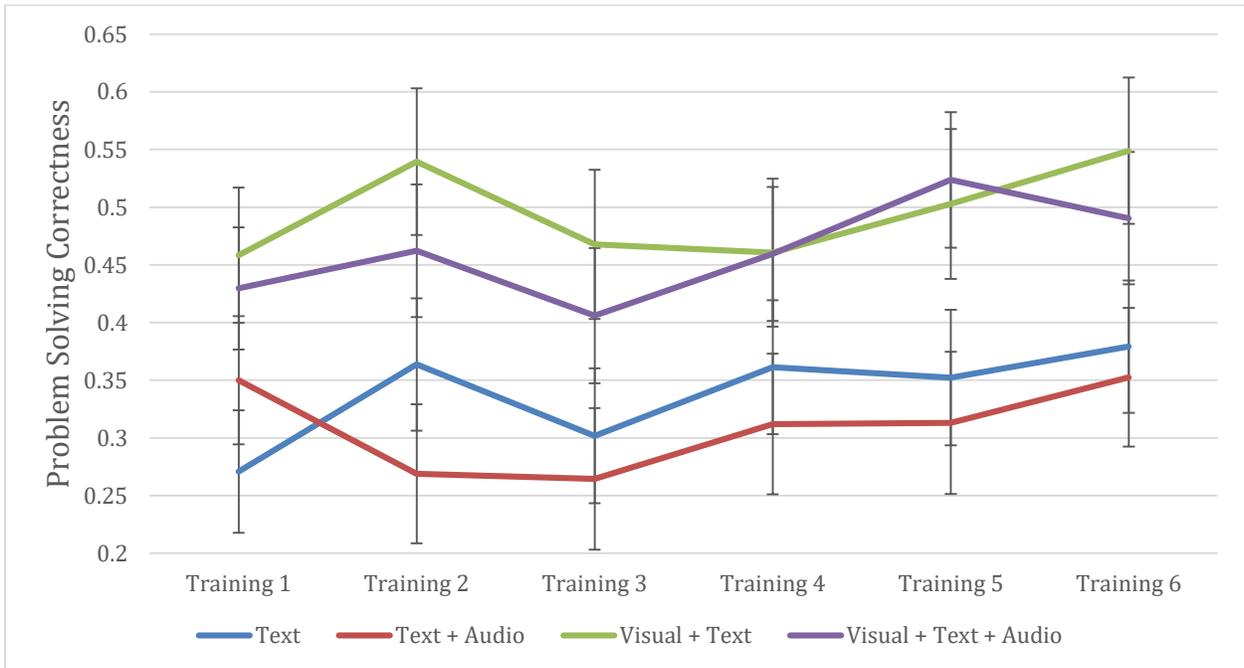
Table 3-3 The _p_ values from pairwise comparisons with Bonferroni correction on the domain relative ratio on text hint AOI between conditions.

Condition	Text	Visual + Text	Text + Audio	Visual + Text + Audio
Text	NA	< .001	1.000	< .001
Visual + Text	< .001	NA	< .001	1.000
Text + Audio	1.000	< .001	NA	< .001
Visual + Text + Audio	< .001	1.000	< .001	NA

Table 3-4 The _p_ values from pairwise comparisons with Bonferroni correction on the domain relative ratio on text hint AOI between training tasks.

Training Task	#1	#2	#3	#4	#5	#6
#1	NA	< .001	< .001	< .001	< .001	< .001
#2	< .001	NA	0.551	0.008	0.003	< .001
#3	< .001	0.551	NA	1.000	0.718	0.001
#4	< .001	0.008	1.000	NA	1.000	0.018
#5	< .001	0.003	0.718	1.000	NA	0.199
#6	< .001	< .001	0.001	0.018	0.199	NA

To probe how participants split visual attention when visual hints and text hints were presented simultaneously, a two-way 2 (visual + text condition, visual + text + audio condition) × 2 (visual hint AOI, text hint AOI) mixed ANOVA with repeated measures on AOI was

conducted (see Figure 3-6). There was no main effect of condition, $F(1, 33) = 0.712$, $p = 0.405$.

It indicated adding audio hints does not affect how participants split their visual attention

between visual hints and text hints. There was a significant main effect of AOI, $F(1, 33) =$

199.59, $p < .001$. The domain relative ratio on visual hints is significantly larger than on text

hints (10.78 ± 0.59 vs. 1.86 ± 0.15, respectively)

To summarize the analyses in this section (see Table 3-8), we found evidence that

presenting visual hints made participants less likely to view text hints and vice versa. This is

consistent with *Hypothesis I*. The comparison between visual hint AOI and text hint AOI showed

participants preferred visual hints to text hints. This result contradicts *Hypothesis II*. Adding

audio hints with visual hints increased the domain relative ratio on the visual hint AOI slightly.

This evidence tends to be aligned with *Hypothesis III-A*. However, there is no significant

difference. Adding audio hints with text hints also increases the domain relative ratio on the text

hints AOI slightly. This evidence tends to be aligned with *Hypothesis IV-A*, but without

statistical significance.

Figure 3-6 The domain relative ratios on the text hint AOI and the visual hint AOI. Error bars indicate ±1 standard error of the mean.
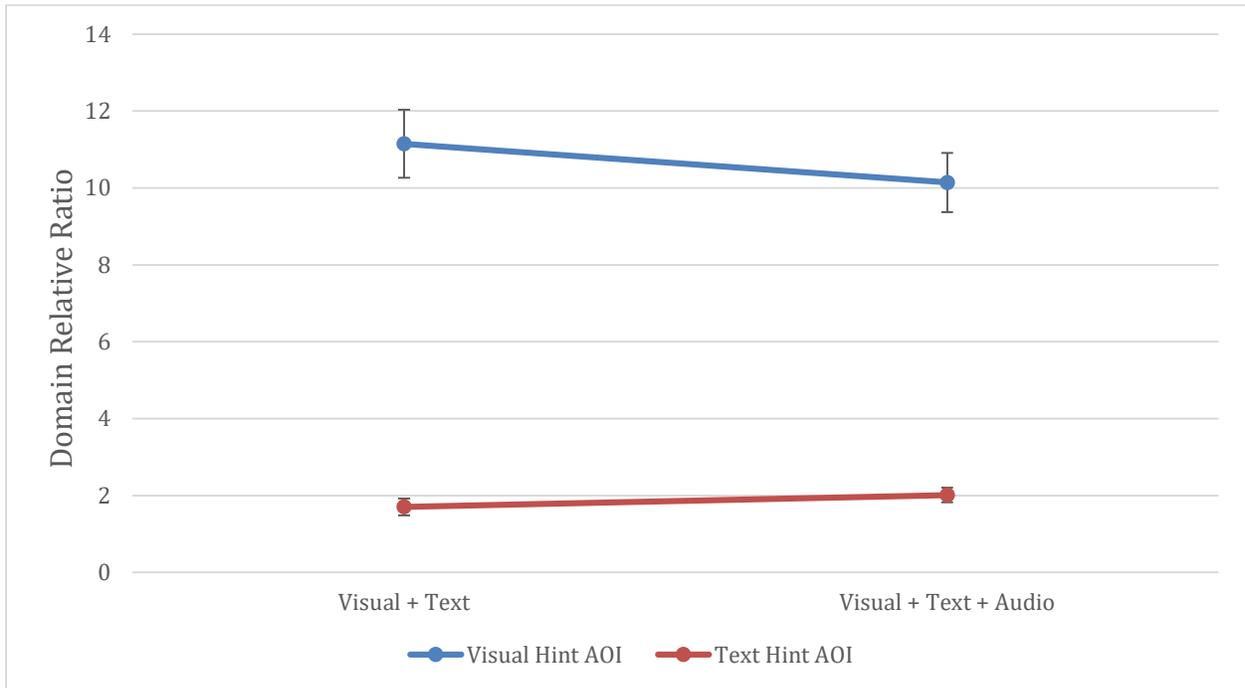
Mean Fixation Duration

The analyses here include fixation durations on the first training tasks with initial task solved incorrectly. A two-way 3 (before hints, during hints, and after hints) × 7 (conditions with hints, *i.e.*, visual condition, text condition, audio condition, visual + text condition, visual + audio condition, and visual + text + audio condition) mixed ANOVA with repeated measures on time intervals (*i.e.*, before hints, during hints, and after hints) was conducted to compare the mean fixation durations. There was a significant main effect of time intervals, $F(2, 962) = 92.03$, $p < .001$. And there was a significant main effect of conditions $F(6, 481) = 7.098$, $p < .001$. These main effects were qualified by a significant interaction, $F(12, 962) = 21.794$, $p < .001$. With probing the interaction, we find that there was significant difference in mean fixation duration before hints in conditions. $F(6, 481) = 2.855$, $p = 0.010$. Since participants were randomly assigned to conditions and they solved the same four task sets, this effect is due to individual differences between participants. To take this effect out of consideration, the

66

ANCOVA with controlling for the mean fixation duration before hints was conducted. A one-way ANCOVA comparing mean fixation duration during hints showed there was a significant effect of condition, $F(6, 430) = 33.638$, $p < .001$. Post hoc tests with the Bonferroni correction were conducted to probe this effect. Table 3-5 presents mean fixation durations before hints, during hints, and after hints for each hint condition. Table 3-6 sums up the statistical results of all pairwise comparisons. Participants in the text condition had significantly shorter mean fixation duration than those in the audio condition. This is consistent with *Hypothesis V*. The mean fixation duration of text + audio condition was slightly longer than the mean fixation duration of text condition. This is consistent with *Hypothesis VI*. But there is no significant difference. Visual + text condition had significantly longer mean fixation duration than text condition. Visual + text + audio condition only had marginally longer mean fixation duration than text + audio condition. These two pieces of evidence together supported *Hypothesis VII-B*. Adding audio hints with visual hints slightly reduced mean fixation duration on visual hints. This result supported *Hypothesis VIII-B* without statistical significance.

A summary of all hypotheses and pairwise comparisons on mean fixation durations are shown in Table 3-9.

Table 3-5 The mean fixation duration before hints, during hints, and after hints (in milliseconds).

Condition	Visual	Text	Audio	V + T	V + A	T + A	V + T + A
Before Hints	245.50 ± 8.33	239.50 ± 7.40	205.77 ± 7.86	241.84 ± 8.26	238.53 ± 8.26	232.17 ± 8.12	248.87 ± 7.20
During Hints	335.79 ± 7.54	218.72 ± 6.69	273.49 ± 7.23	279.49 ± 7.47	316.33 ± 7.47	235.76 ± 7.34	255.67 ± 6.53
After Hints	244.00 ± 6.73	248.99 ± 5.97	268.07 ± 6.44	267.24 ± 6.67	286.92 ± 6.66	259.68 ± 6.55	270.51 ± 5.82

Table 3-6 The *p* values from pairwise comparisons on the mean fixation duration during hints with Bonferroni correction.

Condition	Visual	Text	Audio	V + T	V + A	T + A	V + T + A
Visual	NA	< .0001	< .0001	< .0001	1.000	< .0001	< .0001
Text	< .0001	NA	< .0001	< .0001	< .0001	1.000	0.002
Audio	< .0001	< .0001	NA	1.000	0.001	0.006	1.000
V + T	< .0001	< .0001	1.000	NA	0.011	0.001	0.350
V + A	1.000	< .0001	0.001	0.011	NA	< .0001	< .0001
T + A	< .0001	1.000	0.006	0.001	< .0001	NA	0.915
V + T + A	< .0001	0.002	1.000	0.350	< .0001	0.915	NA

Table 3-7 The *p* values from pairwise comparisons on the mean fixation duration after hints with Bonferroni correction.

Condition	Visual	Text	Audio	V + T	V + A	T + A	V + T + A
Visual	NA	1.000	0.220	0.303	0.000	1.000	0.063
Text	1.000	NA	0.645	0.879	0.001	1.000	0.212
Audio	0.220	0.645	NA	1.000	0.897	1.000	1.000
V + T	0.303	0.879	1.000	NA	0.781	1.000	1.000
V + A	0.000	0.001	0.897	0.781	NA	0.078	1.000
T + A	1.000	1.000	1.000	1.000	0.078	NA	1.000
V + T + A	0.063	0.212	1.000	1.000	1.000	1.000	NA

Table 3-8 Four hypotheses of domain relative ratios and corresponding pairwise comparisons with statistical results

	Pairwise Comparison (AOIs of conditions)	*p*	Consistent with hypothesis?
I	Visual Hint AOI of Visual + Text < Visual Hint AOI of Visual	< .001	Yes
	Text Hint AOI of Visual + Text < Text Hint AOI of Text	< .001	Yes
II	Text Hint AOI of Visual + Text < Visual Hint AOI of Visual + Text	< .001	No
III	Visual Hint AOI of Visual + Audio vs. Visual Hint AOI of Visual	1.000	No difference
IV	Text Hint AOI of Text + Audio vs. Text Hint AOI of Text	1.000	No difference

Table 3-9 Four hypotheses of mean fixation durations and corresponding pairwise comparisons with statistical results

|	Pairwise Comparison (conditions with mean fixation duration during hints)	*p*	Consistent with hypothesis?
V	Text (218.72 ± 6.69ms) vs. Audio (273.49 ± 7.23ms)	< .001	Yes
VI	Text (218.72 ± 6.69ms) vs. Text + Audio (235.76 ± 7.34ms)	1.000	No difference
VII	Visual + Text + Audio (255.67 ± 6.53ms) vs. Text + Audio (235.76 ± 7.34ms)	0.915	No difference
	Visual + Text (279.49 ± 7.47ms) vs. Text (218.72 ± 6.69ms)	< .001	*Hypothesis VII-B*
VIII	Visual (335.79 ± 7.54ms) vs. Visual + Audio (316.33 ± 7.47ms)	1.000	No difference

Discussion

Attention while Hints are Presented

According to Mayer's CTML, the reason that pictorial and textual information cannot make a great combination is presenting two types of visual information would force leaners to split their visual attention. Leaners may miss the important pictorial or textual information. Presenting auditory information with pictorial information does not have this interference. Learners can easily attend auditory information and pictorial information simultaneously with no trouble.

We found evidence to support split attention based on our data analysis. Presenting text hints with visual hints made participants less likely to view visual hints and vice versa. However, the problem solving performance of visual + text condition was significantly better than the visual condition for those who solved the initial problem incorrectly. However, the split attention did not deteriorate problem-solving performance.

A post hoc hypothesis of this situation is that the participants may have sufficient amount of time to read the text hints and view visual hints. To test this hypothesis, we analyzed the numbers of fixations per text hints' content word (see Figure 3-7).

Figure 3-7 The numbers of fixations per text hints' content word. Error bars indicate ±1 standard error of the mean.
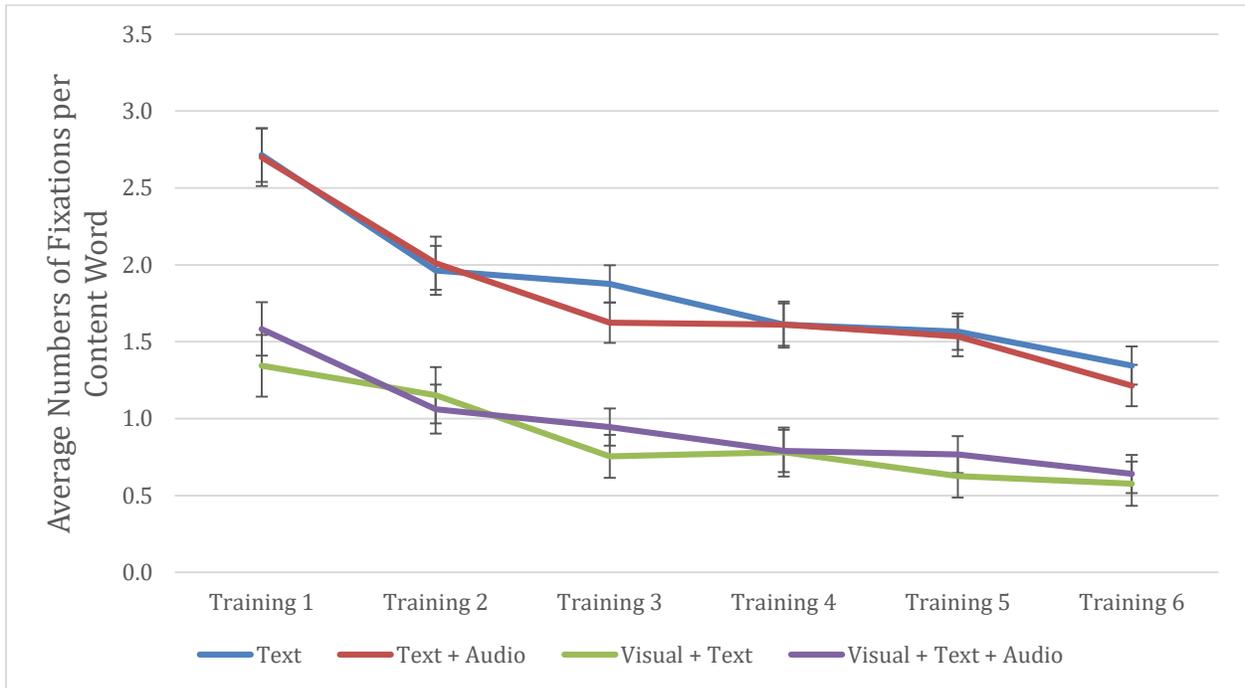
A two-way 4 (conditions with text hints, i.e., text condition, text + audio condition, visual + text condition, and visual + text + audio condition) × 6 (six training tasks) mixed ANOVA with repeated measures on training tasks was conducted to compare the average numbers of fixations per content words of text hints. There was a significant main effect of conditions, $F(3, 68) = 28.813$, $p < .001$ and a significant main effect of training tasks, $F(5, 340) = 69.631$, $p < .001$. There was no significant interaction between conditions and training tasks, $F(15, 340) = 1.683$, $p = 0.074$. Post hoc pairwise comparisons were conducted to probe these two main effects. Table 3-10 and Table 3-11 summarize the statistical results of these analyses.

Table 3-10 The *p* values from pairwise comparisons with Bonferroni correction on the average numbers of fixations per content word of text hints between conditions.

Condition	Text	Visual + Text	Text + Audio	Visual + Text + Audio
Text	NA	$< .001$	1.000	$< .001$
Visual + Text	$< .001$	NA	$< .001$	1.000
Text + Audio	1.000	$< .001$	NA	$< .001$
Visual + Text + Audio	$< .001$	1.000	$< .001$	NA

Table 3-11 The *p* values from pairwise comparisons with Bonferroni correction on the average numbers of fixations per content word of text hints between training tasks.

Training Task	#1	#2	#3	#4	#5	#6
#1	NA	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$
#2	$< .001$	NA	0.002	$< .001$	$< .001$	$< .001$
#3	$< .001$	0.002	NA	1.000	0.047	$< .001$
#4	$< .001$	$< .0001$	1.000	NA	1.000	$< .001$
#5	$< .001$	$< .001$	0.047	1.000	NA	0.022
#6	$< .001$	$< .001$	$< .001$	$< .001$	0.022	NA

Three important features emerge from these statistical analyses. First, all four conditions' average numbers of fixations were larger than one on the first training tasks. This evidence suggests participants read text hints fully the first time they saw them. Second, all four

conditions showed a decreasing trend. It suggests that participants did not read text hints very carefully on the later training tasks. This is due to the fact that text hints in each task set are identical from the first training task to the last training task. Participants just skimmed the text hints on those later training tasks to know this piece of information was the same as the information from the earlier tasks. Third, adding audio hints with text hints did not affect participants' reading behavior while adding visual hints changed their reading behavior. This evidence confirms the split attention we found from the analyses of the domain relative ratios.

Overall, the post hoc hypothesis is supported by the results of statistical analyses. Participants had sufficient time to read text hints. Compared to visual condition and text condition, participants in visual + text condition could attend to two pieces of help information on solving each training task. This could be the reason that visual + text condition has the better problem solving performance than visual condition and text condition.

On the other hand, we found that participants preferred visual hints to text hints. This evidence contradicts *Hypothesis II*. Suggestive explanations could be proposed on two aspects: the pace of hints and the visual design of hints.

Many studies on multimedia instruction found the instruction pace was a critical factor influencing learning outcomes (Harskamp, *et. al*., 2007; Stiller, *et. al*., 2009) and visual attention (Schmidt-Weigand, *et. al*., 2010b). It has been found that slow instruction pace would extend leaners' viewing time on visualizations more than viewing time on the text. In Schmidt-Weigand's eye tracking study (2010a), the slow pace was set to be a reading rate of 67.5 words per minute and the medium pace was 90 words per minute. As for speaking speed, Blau's (1990) study on speaking speed set approximately 170 words per minute as the normal speed and approximately 145 words per minute as the slow speed. Our text hints contained about ten words

for each problem and the duration of hints was set as eight seconds, which corresponds to about 75 words per minute. It made our hints fit the slow-pace category, in which leaners were willing to spend the extra time on viewing elements of the visualization.

However, pace of hints cannot explain our participants' preference on visual hints alone. With the slow pace instruction, learners' from Schimidt-Weigand's study (2010a) spent about equal amount of time on viewing text and visualization. This is also consistent with Rayner's study (2001) on viewing print advertisements. A close look at the material used in Schimidt-Weigand's multimedia instruction study and the print advertisements used in Rayner's study reveals that they share many similarities. Textual statements in these two studies were all in the central position to attract participants' attention, and the textual information is critical for viewers to understand the material. The multimedia instruction module in Schimidt-Weigand's (2010a) study explains the formation of lightning. Even though the part of the visualization shows the whole event, it would be difficult to understand the mechanism of lightning formation without reading the text. As for the advertisements in Rayner's study, they are just pictures of a car and a woman's profile without the superimposed text. The fact that learners always attempt to integrate pieces of external information to build a mental model would force viewers to read the text to have a better understanding. In our study, text hints were put close to the diagram and visual hints according to Mayer's *spatial contiguity principle* (2005b). The visual hints in bright yellow and flashing manner increased their visual salience to overpower the text hints. Further, visual hints by themselves can help participants solve tasks significantly better than non-visual hints. Participants in visual conditions solved tasks significantly better than the no hint condition (see Chapter 2). Visual hints do not need to be presented with text hints to be explained to participants. So participants in our study preferred to look at visual hints.

As for audio hints, two hypotheses testing the effect of adding audio hints did not have significant effects. Moreover, testing these two pairs of competing hypotheses indicated that Wickens' (2002) MRT cannot provide a coherent explanation toward participants' eye movements in our study. In regard to either spatial information or linguistic information, adding audio information always increased participants' viewing/reading time. This evidence will be further discussed in a later paragraph with the data of mean fixation durations.

This piece of evidence of splitting attention could help us understand why we found adding text hints improved visual hints. It is true that presenting text hints with visual hints would draw participants' visual attention away from visual hints. But the improved problem solving performance showed that the amount of time participants spent on reading text hints was worthwhile. Especially for those who did not know how to solve the initial tasks, reading text hints gave them the edge they needed.

To see how multimedia hints affected participants' visual attention after hints, we studied the domain relative ratios of novice AOI before hints and after hints, and the domain relative ratios of expert AOI before hints and after hints (see Figure 3-8). The definitions of novice AOI and expert AOI were the same as our group's previous studies (Rouinfar, 2014; Agra, 2015, see Figure 3-9 for example). We used the domain relative ratios of novice AOI before hints divided by the domain relative ratios of expert AOI before hints. We did the same calculation for the domain relative ratios after hints. Then a new variable, which we call "expert index", which is the ratio of domain relative rations after hints to before hints (no hint condition had no "after hints". Therefore the value of this condition was designated to be 1), was calculated to conduct a one-way ANOVA to test the effect of hints after hint presented. Table 3-12 shows the means and standard errors of "expert index" in each condition along with the mean and standard errors of

the percentage correct rate in each condition. Figure 3-10 shows a graph of the 'expert index' for each condition.

We found a significant effect of condition, $F(7, 535) = 3.795$, $p < .001$. The post hoc pairwise comparisons with Bonferroni correction found visual + audio condition had the highest expert index and audio condition had the marginally lowest expert index (see Table 3-13). This "expert index" reflected the extent of participant shifting visual attention from novice AOI to expert AOI after hint presented. Therefore, audio hints were the least effective in changing participants' visual attention and visual + audio hints were the most effective on training participants to attend to the expert areas of diagrams. Due to the fact that audio condition had the worst performance among hint conditions and visual + audio condition has the second best performance, the extent of shifting visual attention after hints aligns with the problem solving performance well.

Figure 3-8 The domain relative ratios on Expert AOI and Novice AOI before hints and after hints in each condition.
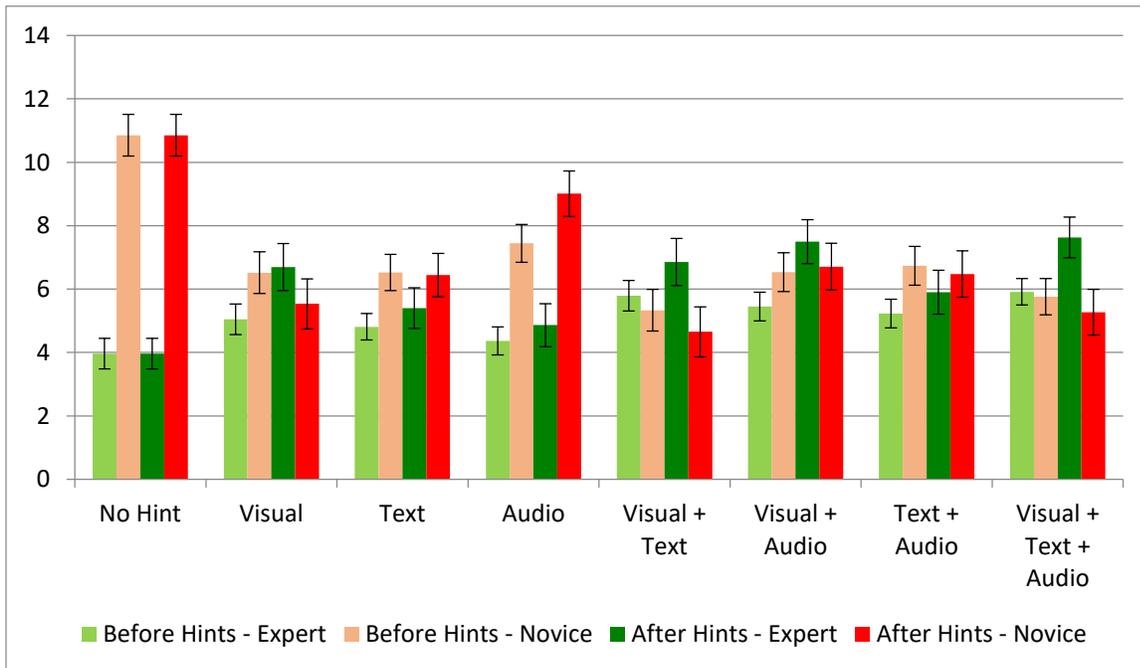
Figure 3-9 An example of drawing expert AOI and novice AOI in a slide of a ball task.
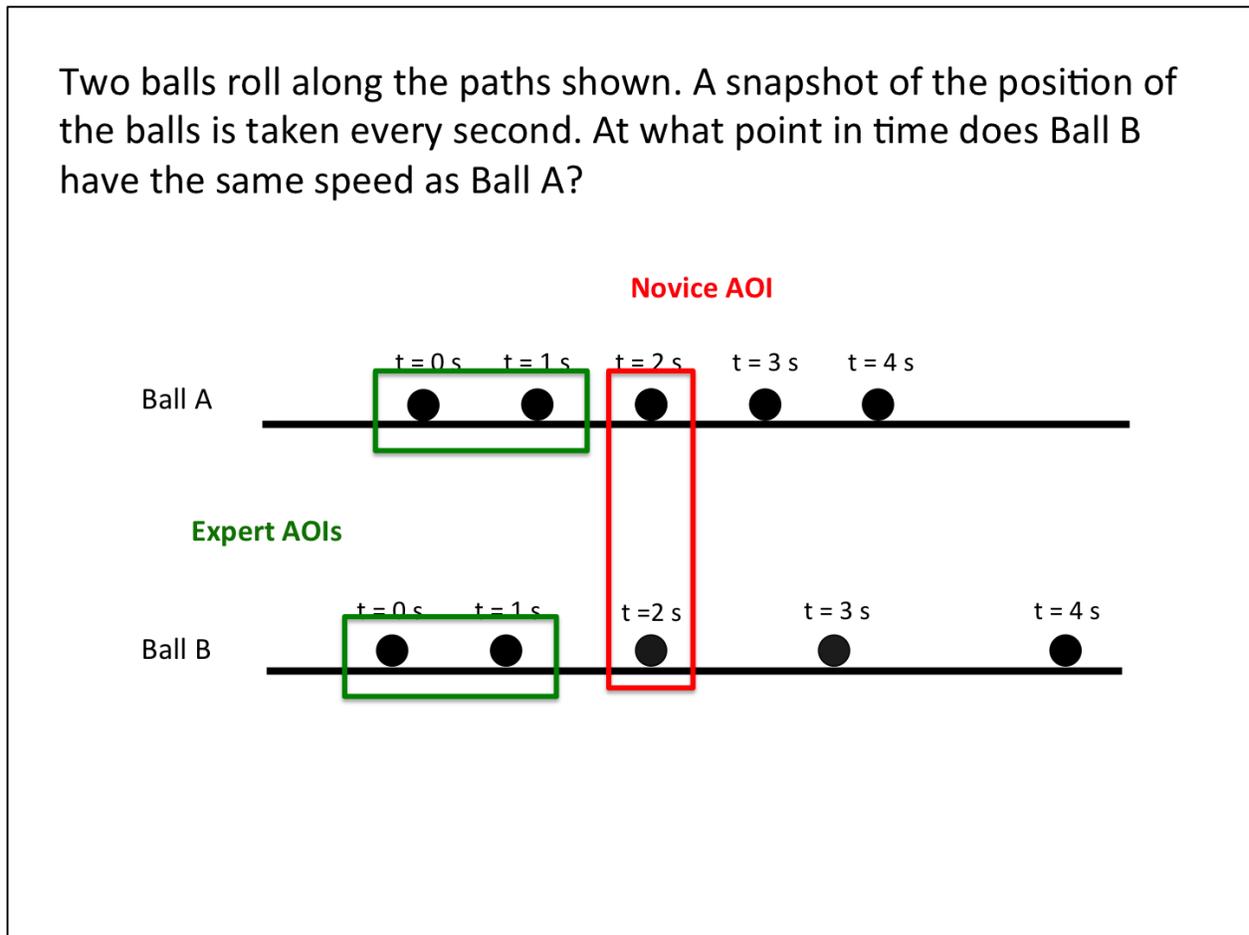
Table 3-12 The means and standard errors of expert index and the means and standard errors of percentage of correct rate in each condition.

Training Task	Expert Index	SD of Expert Index	Percentage of correct rate	SD of Percentage of correct rate
No Hint	1	0.00	12.8%	4.9%
Visual	3.00	0.63	34.8%	4.9%
Text	2.03	0.55	33.0%	4.2%
Audio	1.51	0.58	24.9%	4.5%

Training Task	Expert Index	SD of Expert Index	Percentage of correct rate	SD of Percentage of correct rate
Visual + Text	2.29	0.63	43.0%	4.7%
Visual + Audio	4.80	0.60	41.3%	4.6%
Text + Audio	1.70	0.59	30.0%	4.7%
Visual + Text + Audio	2.76	0.55	39.7%	4.2%

Figure 3-10 The expert index in each condition.
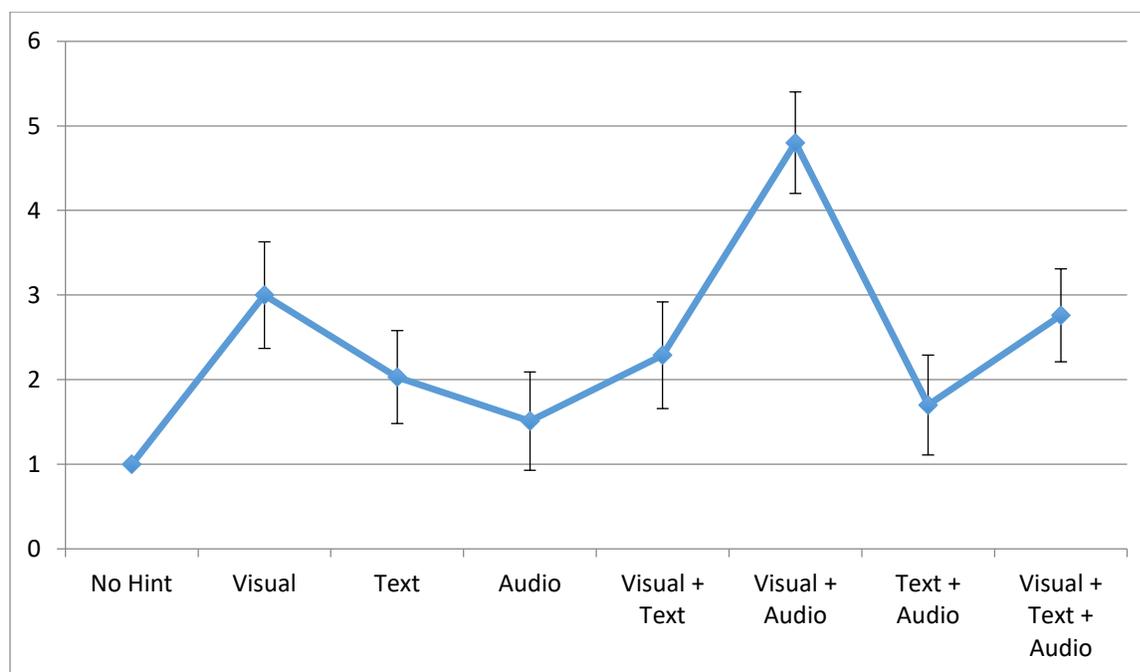
Table 3-13 The *p* values from pairwise comparisons with Bonferroni correction on the Expert Index between conditions.

Training Task	No Hint	Visual	Text	Audio	Visual + Text	Visual + Audio	Text + Audio	Visual + Text + Audio
No Hint	NA	0.698	1.000	1.000	1.000	< .001	1.000	0.980
Visual	0.698	NA	1.000	1.000	1.000	1.000	1.000	1.000
Text	1.000	1.000	NA	1.000	1.000	0.020	1.000	1.000
Audio	1.000	1.000	1.000	NA	1.000	0.002	1.000	1.000
Visual + Text	1.000	1.000	1.000	1.000	NA	0.112	1.000	1.000
Visual + Audio	< .001	1.000	0.020	0.002	0.112	NA	0.007	0.338
Text + Audio	1.000	1.000	1.000	1.000	1.000	0.007	NA	1.000
Visual + Text + Audio	0.980	1.000	1.000	1.000	1.000	0.338	1.000	NA

Cognitive Load during Problem Solving

We used mean fixation duration to assess cognitive load in our study. The results were not all consistent with our hypotheses. The conditions with visual hints had longer mean fixation durations than the conditions without visual hints, and text hints seemed to cause less cognitive load than audio hints.

The three time intervals in problem solving: before hints, during hints, and after hints were designed to allow participants to make step-wise progress in each time interval. The time interval of before hints was for participants to read and understand the task. The time interval of during hints was for them to perceive hints. And participants were supposed to construct a response and report their answers in the time interval of after hints. Since participants had different objectives in these three time intervals, it is necessary to put mean fixation durations before hints, during hints, and after hints for all hint conditions together to see what effect visual hints, text hints, and audio hints had on our participants.

Figure 3-11 shows the mean fixation duration ratios. The ratios were calculated by comparing mean fixation durations before hints, during hints, and after hints with the mean fixation duration before hints. Among these line plots, the striking features of the line of visual hints and the line of text hints draw our attention.

Figure 3-11 The ratios of mean fixation durations on all seven conditions with hints. Error bars indicate ±1 standard error of the mean.
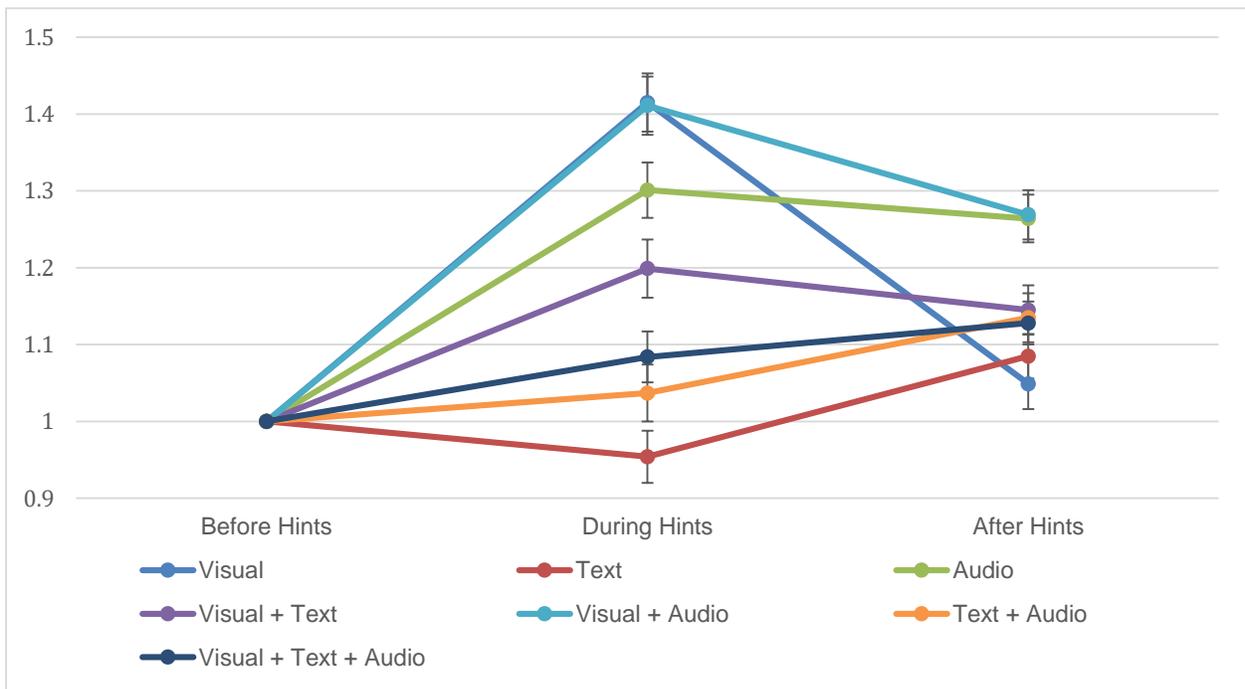
Visual condition was the condition with the longer mean fixation duration during hints. The explanation we suggest here is the difference between the visual condition and text condition on the mean fixation durations is mainly due to the difference between reading and scene perception (Rayner, 1998). According to Rayner's review, the ratio between the mean fixation duration in scene perception, and the mean fixation duration in reading is 1.47. In our study, the ratio between the mean fixation duration during hints in the visual condition and the mean fixation duration during hints in the text condition is around 1.48, which is very close to the number reported by Rayner. Moreover, the mean fixation durations during hints of visual + text condition and visual + text + audio condition are shorter than the visual condition's and longer than the text condition's. This also suggests that participants mixed reading text hints and inspecting visual hints in these two conditions.

We also found that the mean fixation duration during hints in the text condition was significantly shorter than the mean fixation duration during hints in the audio condition. This evidence suggests that participants in the audio condition had more mental effort on understanding the information from auditory modality than those in the text condition. The performance data analyses also showed the text condition was significantly better than the audio condition. Therefore, participants' difficulties on perceiving audio hints might be the reason for this reverse modality effect. In their study on reverse modality effect, Leahy and Sweller (2011) found that students preferred visual text to audio text when the information was complex. They suggested the information with multiple elements and high element interactivity would require much mental effort and was better presented in written form. Due to its transient nature, audio information is not easy for learners to understand and integrate with previously presented

information. However, the text information is permanent during the presenting time. Learners can easily go back and forth to relate all pieces of information.

Table 3-14 lists all linguistic hints and the number of relevant elements. The contents of the hints were all asking participants to compare multiple elements. It is clear that each hint contains complex information for participants to attend to. Therefore, the complexity of the linguistic information made it easier for participants to read text hints more than listen to audio hints because with text hints presented on screen, they did not need to hold every piece of the information presented in their working memory for post processing after the hints were no longer present, i.e. in the after hint situation.

Table 3-14 The linguistic hints and numbers of relevant elements on each task set.

Task	Linguistic Hints	Numbers of Interacting Elements
Skier	Compare the change in height for each section.	4 sections
Ball	Compare the distance between subsequent snapshots of the two balls	8 distances
Roller Coaster	Compare the change in position from initial to final for two carts.	4 positions
Graph	Compare the slope along the curve at each moment in time	4 slopes

As for audio hints, we found no significant effect on both adding audio hints with text hints and adding audio hints with visual hints. Adding audio hints with text hints would slightly increase the mean fixation duration during hints. This is still true when the mean fixation duration before hints is controlled. While the small difference between visual + audio condition and visual condition disappears when the mean fixation duration before hints is controlled. They are almost identical (see Figure 3-11).

The aforementioned evidence enables us to propose a hypothetical explanation: participants in our study tried to ignore audio hints when audio hints were presented with other hints. The small differences, if there were any, between conditions with audio hints and conditions without audio hints (excluding the no hint condition) on the mean fixation duration are due to the mental effort of blocking out audio hints. This is why we found participants viewed/read visual/text hints for slightly longer time when there were audio hints and participants had slightly longer mean fixation duration on reading text hints when there were audio hints. This suggestive explanation is not consistent with unpublished work done by Sohl (1989), which showed evidences that participants did try to follow the speech when speech and subtitles were played simultaneously. But this explanation is a good fit with our eye movement data. The only exception is when adding audio hints with visual + text hints decreased mean fixation durations. But consulting with the corresponding domain relative ratios (see the plots of visual + text condition and visual + text + audio conditions in Figure 3-2 and Figure 3-4) would tell us this is due to adding audio hints shifts participants' attention slightly toward text hints and reading has shorter mean fixation duration than scene perception.

We mainly focused on the mean fixation durations during hints to discuss the cognitive load imposed by hints until now. We found text hints imposed less cognitive load than audio hints. But due to the different visual stimuli, we cannot use mean fixation duration as the indicator of cognitive load for the comparison between visual condition and text condition, but we can compare the mean fixation durations of visual hints before hints, during hints and after hints to explore the change pattern. This pattern could reflect the cognitive load change for presenting visual hints in the whole problem solving procedure. A one-way ANOVA with repeated measures on time intervals (i.e., before hints, during hints, and after hints) was

83

conducted to compare the mean fixation durations of the visual condition. There was a significant main effect of time interval, $F(2, 118) = 70.135$, $p < .001$. Post hoc tests using the Bonferroni correction revealed that the mean fixation duration during hints was significantly larger than the mean fixation duration before/after hints (333.09 ± 10.77ms vs. 240.87 ± 8.65ms/245.09 ± 5.21ms, respectively). There was no significant difference between the mean fixation durations before hints and after hints (240.87 ± 8.65ms vs. 245.09 ± 5.21ms, respectively). These tests showed that presenting visual hints increased cognitive load while the cognitive loads before hints and after hints were at about the same level. This result will be further explained in Chapter 4.

In summary, the eye movement data helped to explain the problem solving performance with multimedia hints. The duration of hints in our study was set to be sufficiently long for reading text hints and inspecting visual hints. Therefore, the split attention effect did not make participants miss the critical information from visual hints and text hints. The extra linguistic help from text hints boosts visual + text condition's performance. Our problem solving performance with split attention is not a contradiction of Mayer's CTML due to the amount of time we provided student to attend to hints. As for the other linguistic modality, audio hints did not help participants as much as text hints. Attending to audio hints would cause more cognitive load than reading text hints. Moreover, we suggest that participants may try to ignore audio hints when audio hints were presented with other hints simultaneously. Overall, audio hints by themselves were not an effective way to convey the help information and were not effective when accompanied with other hint modalities either.

Chapter 4 - A Model of Physics Problem Solving with Multimedia Hints

Introduction

In this chapter, we propose a conceptual model that describes physics problem solving with multimedia hints. This model integrates concepts from Ohlsson's (1992) Representational Change Theory (RCT), Mayer's (2002) Cognitive Theory of Multimedia Learning (CTML), Wickens' (2002) Multiple Resources Theory (MRT), and the framework of attention cueing (de Koning, *et al.*, 2009). The model also attends to the issues of cognitive load theory (Sweller, 1988) and working memory (Baddeley, 1992). This model seeks to provide a coherent explanation for the data presented in Chapter 2 and Chapter 3.

Theoretical Background

Theories of Problem Solving

The theory guiding our experimental design was the Cognitive Theory of Multimedia Learning -- CTML (Mayer, 2002). This theory has many successful applications in multimedia learning. However, task performance and learning are fundamentally different (Schnotz and Kürschner, 2007). Learning is making changes in long-term memory while performing a task is a cognitive process to alter mental representation in working memory. Learners may practice on some learning tasks to facilitate learning. But solving tasks may not directly influence learner's long-term memory and result in learning. For example, to learn Faraday's Law, learners may be asked to solve for current on the moving rod in a magnetic field, or the electric potential on the spinning rod in a magnetic field. To solve those problems, learners need to transform the given state (i.e., moving or spinning rod in a magnetic field) into the goal state (i.e., the current or

electric potential on the rod). Understanding Faraday's Law, which is only a byproduct of this problem-solving activity, is not necessarily triggered in learners' long-term memory when learners are engaging in solving those problems. The scope of our study is mainly on conceptual physics problem solving. Therefore, the pair of questions we should answer is, what is a problem and what is problem solving?

In their early work, Newell and Simon (1972) defined a problem as the situation that a person wants something and does not know immediately how to get it. They adopted the view of a human information processing system to model human problem solving. The original theory they proposed has been long discussed and developed, but the propositions they outlined to shape their problem solving theory can still shed light on theoretical problem solving studies. They argued that given the large diversity of problem solvers and tasks, looking for the invariants over tasks and solvers, should be the foundation of a problem solving theory.

Many cognitive psychologists started from investigating a small set of special tasks: insight problems. The traits of insight problems have been well documented. Dow and Mayer (2004) provided a practical definition of an insight problem: "…a special type of non-routine problem in which the problem primes an inappropriate solution procedure that is familiar to the problem solver" (pp. 389). This descriptive definition fits the nature of many well-studied insight problems. The concepts of insight and impasse, as a pair, cannot be avoided in the discussion of insight problems. Empirically, an insight is always connected with the "aha" moment: "Now I see how to solve this problem." The "aha moment" is the moment of breaking impasse. Regardless of the reason of impasse, it is a common feature of insight problems.

The particularity of insights and insight problems makes the discussion of this type of problem solving different from the traditional information-processing view. The reconciliation

86

started from Ohlsson's modified RCT (Ohlsson, 1992). In his chapter, Ohlsson, as a researcher advocating insights and insight problem solving for a long time, explained insight problem solving through the lens of information processing. Unlike the classical definition, which puts an insight as the sudden appearance of a complete and correct solution (Ohlsson, 1984), Ohlsson re-defined insights as "initial failure followed by eventual success" (pp. 5). The initial failure is the impasse that the capable solver cannot solve the problem. The momentary impasse breaking is the reason of the subjective sudden feeling. An insight is not necessarily complete or correct. Multiple insights might be needed to solve a complex problem. A partial insight might occur while solving a problem stepwise. An insight could also be wrong and lead the solver to a new impasse.

Modified RCT provides an explanation for impasse generating and breaking. A capable solver, who in principle possesses all the necessary knowledge to solve the problem, is facing an impasse because the mental representation of the problem in the solver's mind limits the activation of necessary knowledge. To break the impasse, the incorrect mental representation needs to be altered. Ohlsson (1992) suggested three mechanisms for impasse breaking: elaboration, re-encoding and constraint relaxation. Elaboration is adding extra information internally (recalling information from long-term memory) or externally (re-studying the problem situation or receiving hints). Re-encoding is rejecting the old problem representation and constructing a different representation. Constraint relaxation is removing the constraints on the ways of reaching the goal situation. After an impasse is broken, problem solving resumes, and the solver can progress toward the goal of the problem via the newly discovered path.

Cognitive Theory of Multimedia Learning (CTML)

Figure 4-1 presents how external information proceeds through the human information-processing system based on Mayer's CTML (2002).

Figure 4-1 Cognitive theory of multimedia learning (Mayer, 2002)
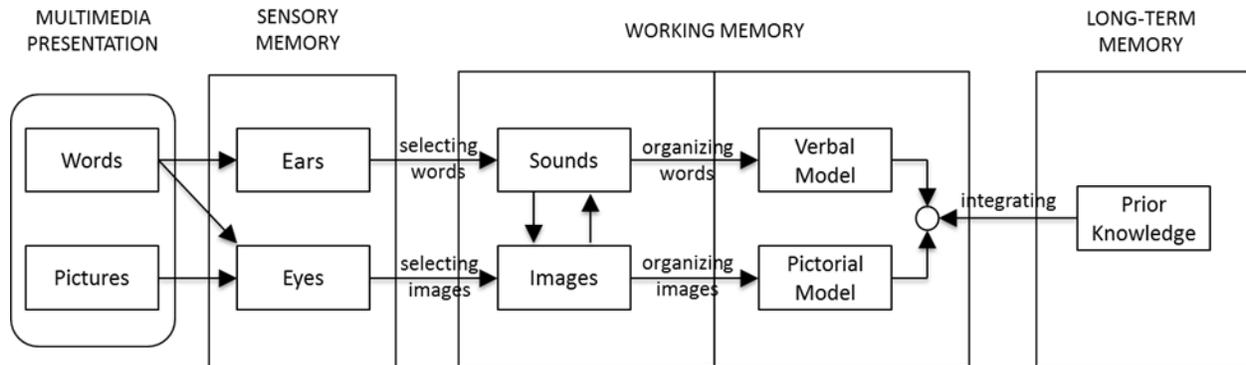
The memory stores have three levels. They are sensory memory, working memory, and long-term memory. The external information is perceived by the ears and eyes and temporarily stored in the sensory memory. Then the relevant words and images in the sensory memory are selected, and sent into working memory. Working memory is the place to actively store and process information (the function and construction of working memory will be reviewed in the later section). In working memory, the information of words is organized into a verbal model, and the information of images is organized into a pictorial model. To understand the external information, the verbal and pictorial models need to be integrated. The information from two different representations is connected into a coherent representation. This coherent representation also includes the activated prior knowledge from long-term memory.

This information processing system has two channels and they are separated by *sensory modality*, i.e., the ears and the eyes. The upper pathway in Figure 4-1 presents the auditory channel and the lower one is the visual channel. When the external information is spoken words,

they are perceived by the ears and stored in auditory sensory memory. Then the relevant auditory information is selected and organized into the verbal model. When the external information is pictures, the steps are similar but located in the visual channel. Processing printed words is an interesting situation. According to Mayer's CTML (2002), printed words are picked up by the eyes and selected as images. Next, the images are transferred into sounds in working memory. The arrows between the box of sounds and the box of images represent this process.

Multiple Resources Theory

Mayer's CTML discussed the mechanism of processing external information. When students are engaging with a task, processing external information is only one factor of task performance. Wickens' (2002) Multiple Resource Theory (MRT) is the theory to fully discuss all aspects of performing tasks. Figure 4-2 represents the four-dimensional model proposed by MRT. Each dimension has two levels.

Figure 4-2 Four-dimensional model of multiple resources. The fourth dimension (visual processing) is within the dimension of visual resources (Wickens, 2002).
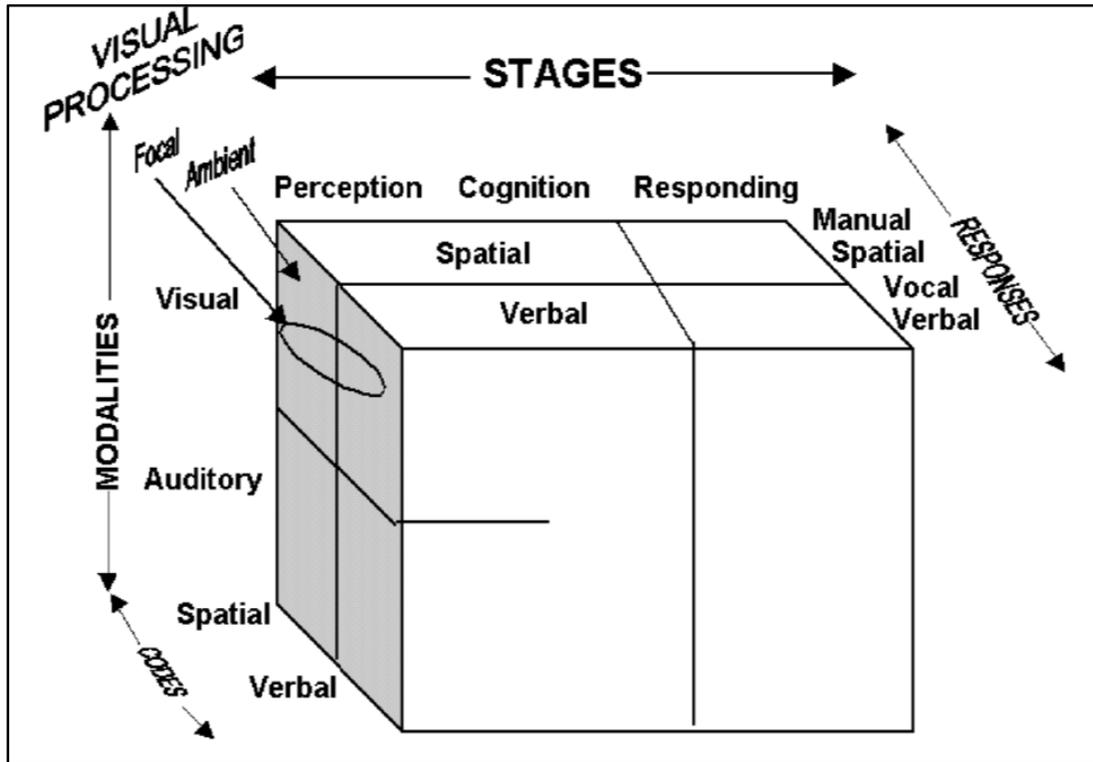
If two tasks are equal in all aspects and share a process resource at one level, then engaging in these two tasks in parallel will cause interference and damage performance for both of the tasks. The first dimension is stages. There are two levels. Perception and Cognition share one level and responding is another level. This is supported by physiological evidence (Israel, *et al.*, 1979). The second dimension is modalities. This dimension contains visual channel and auditory channel. This part of the model is consistent with Mayer's CTML. The third dimension is visual channels. This dimension further divides visual channels into two levels: focal and ambient. This separation is supported by their different types of information processing (Weinstein and Wickens, 1992). The fourth dimension, codes, is built on the distinction between linguistic and spatial processes.

The most important application of this theory is to predict the performance of multiple time-shared tasks. For example, Goodman *et al.* (1999) suggested that when comparing manual dialing of cell phones during driving, with voice dialing, voice dialing has a smaller effect on driving performance. This is because driving uses the resources of visual (one level of modalities) and spatial (one level of codes) and manual dialing also uses the resources of visual and spatial. However, voice dialing uses the resources of linguistic and auditory. They are not the same level as the resources used by driving. Therefore, voice dialing is safer than manual dialing during driving.

The Effects of Hints

Hints are widely used with different formats and different functions. Instructors use hints to prompt students to think deeply on concepts. Tutors use hints to guide students to move to the next step during problem solving. Books use headers or fonts to help readers organize information, and television programs use annotations to draw the attention of their audience. There are many different types of hints and they have many different effects on perception and cognition. This section mainly adopts the framework of attention cueing (de Koning, *et al.*, 2009) to discuss the formats and functions of hints.

The attention cueing framework summarizes three hint functions: selection, organization, and integration. These three functions are grounded in the discussion of CTML (Mayer, 2002). According to this theory, the external information needs to be selected, organized, and then integrated by learners to facilitate learning. Therefore, these three steps of information processing categorize the functions of hints. It must be noted that even though the goal of this attention cueing framework is to exam the possibility of transferring different types of hints into animation, the hints included there are both visual and linguistic.

Selection is mainly the function of visual hints. This kind of hints has been used in support text comprehension (de Koning *et al*., 2009). Since it is not always the case that readers are familiar with the background knowledge of a book, differentiating relevant information and irrelevant information could be difficult (Bromage and Mayer, 1981). This difficulty might result in misunderstanding the main topic of the content. Therefore, helping readers to select relevant information is important. Bold texts and underlining words are simple techniques to tell the reader that there is something important there. These techniques are also used in many computer-based multimedia instructions.

So far we have discussed selecting hints on linguistic information. Selecting hints could also work with pictorial information. Thomas and Lleras' study (2007) conducted a study on providing hints to help problem solving. They used letter/digit sequence to guide participants' visual attentions on the pictures of Duncker's radiation problem (1945). They found participants with this kind of visual hint solved the problem better than participants without visual hints. Though the participants considered the letter/digit sequence as an unrelated task. In this study, the letter/digit sequence served as a device to attract participants' visual attention toward the areas that were related to the correct solution. It helped them select a specific part of the picture to focus on.

The function of selection could work on both pictorial, and linguistic information. And both linguistic hints and visual hints could have this function. The studies we discussed above all used visual hints to help participants do selection. Thomas and Lleras used letter/digit sequence in their study, but the purpose of presenting the letters or digits was not for participants to understand the content, but to move their eyes to the appropriate places. Linguistic hints have also been used in many psychological studies to guide visual attention to the designated areas or

directions. A study of controlling visual attention (Hommel, *et al.*, 2001) compared the

effectiveness of directing visual attention with symbols (i.e., icons of arrow), and direction words

(i.e., up, down, left, and right). Ho and Spence (2005) have studied the effect of verbal words on

drivers' visual attention.

The second function of hints focuses on organization. Visual information may contain

many different elements. Some of them are related with each other and some of them are not.

Organizing hints could help learners understand the organization of these learning materials. For

example, in Lorch *et al.*'s study (1993), participants showed better retention on content with

headings, overviews, and summaries. These were all linguistic hints emphasizing the

organization of linguistic information. It could be expected that providing linguistic hints could

also reveal the organization of pictorial information. For example, the study conducted by Grant

and Spivey (2003) used verbal hints to organize the information from the diagram and improved

them problem-solving performance, but there is limited evidence to show that visual hints can

organize information. Mautone and Mayer (2007) studied how students learn from a multimedia

lesson with built-in graphic organizers. They found that students generated more relational

statements in summarizing the content of the lesson afterward, but graphic organizers are extra

pieces of complete information with both pictorial, and linguistic information, and serve as an

introduction or guidance of the whole lesson. The reason for the limited effectiveness of visual

hints on organizing has been suggested by the paper proposing attention cue framework (de

Koning, *et al.*, 2009). It argued that visual hints have the limitation of accurately presenting

complex information, which might be needed for organizing information.

The third function is integration. Unlike emphasizing organization, integration is

highlighting the relationship between elements. Color-coding is a popular technique to connect

elements together. This connection could be between or within visual and verbal elements. Kalyuga *et al*. (1999) found color-coding relating the words in paragraphs and visual elements in diagrams improved the retention for coded content. Our group's previous study (Rouinfar, 2014) also showed that highlighting multiple visual elements sequentially could trigger participants to compare those elements. Clearly, color-coding is not the only type of hint that could facilitate integration. Brief linguistic information could also serve as integrating hints. Corkill (1992) reviewed 30 experiments on using advanced textual summaries. Most of the experiments demonstrated improving retention. According to *cognitive load theory*, learning could be improved by physically connecting two relevant sources of information to reduce the cognitive load of mental integration (Mayer and Moreno, 2003). Integrating hints could benefit learning by explicitly emphasizing specific relationships between relevant elements.

Model of Working Memory

Research has shown that problem solving relies heavily on working memory (Stevenson and Carlson, 2003; Conway *et al*., 2005). As discussed in Ohlsson's RCT (1992), problem solving starts from extracting problem information. In most of the cases, the problem information is visually or verbally presented to solvers, and processed by working memory. Ericsson and Delaney (1999) suggested "working memory is so central to human cognition that it is hard to find activities where it is not involved" (pp. 259). Due to the critical role that working memory plays in problem solving, we must carefully explore working memory to better understand the process of problem solving. Baddeley's model of working memory (Baddeley, 1992) provided us a detailed map of working memory.

Memory is not a single unitary system. The studies of brain-damaged patients provided the most convincing evidences for a dichotomy. Studies on patients with amnesic syndrome,

found these patients showed the limited long-term learning performance, but normal short-term memory span (Vallar and Papagno, 2002). Atkinson and Shiffrin (1968) first proposed a model of memory with short-term memory and long-term memory. They suggested short-term memory as one place temporarily storing information plus facilitating relevant cognitive activities. To better describe the function of short-term memory, Baddeley and Hitch (1974) proposed a model of working memory, and this model has replaced the old short-term memory model (Crowder, 1982). Originally, the working memory model contained three parts: the *central executive*, the *visuospatial sketchpad*, and the *phonological loop*. The *visuospatial sketchpad* actively stores visual information, and the *phonological loop* actively stores audio information. The *phonological loop* has two components: the phonological store, which holds the acoustic traces for a matter of seconds; and the articulatory rehearsal system, which maintains the acoustic traces by sub-vocal repetition. These two components work together as a loop to register linguistic information. This description is also supported by many read comprehension studies done by Richardson and his colleagues (Richardson, 1987; 1996; Richardson, *et al.,* 1996). As for the *visuospatial sketchpad*, neuropsychological studies suggested the distinctions between visual memory and spatial memory (Della Sala, *et al*., 1999). Visual memory is associated with patterns, and spatial memory is associated with locations. The experimental results from the tasks on visual and spatial memory demonstrated the *visuospatial sketchpad* processing of the visual and spatial information separately. Unlike the clear definition of the articulatory rehearsal system of the *phonological loop*, the visuospatial rehearsal is a bit controversial (Logie, 1995; 2011). Moreover, the separation and definition of the subsystems of the *visuospatial sketchpad* have more difficulties than the dissection of the *phonological loop* (Della Sala and Logie, 2002). One interesting implementation of working memory, with visual-oriented and audio-oriented

components, is interpreting the linguistic information process. Linguistic information could be conveyed by either visual text or audio text. If audio linguistic information is stored and maintained in the phonological loop, then how would the *visuospatial sketchpad* deal with the visual linguistic information? Baddeley (2003) suggested that the linguistic information from visual input would first be analyzed and temporally stored in visual short-term storage, and then be recoded from orthographic information to phonological information. According to this idea, visual linguistic information is mainly dealt by the *phonological loop* once it has been re-encoded. Therefore, the *visuospatial sketchpad* is responsible for pictorial information, while the *phonological loop* is responsible for linguistic information.

The third subsystem of working memory is the *central executive*. The main function of this subsystem is to coordinate information from different sources and control attention. Since the *central executive* is responsible to distribute resources to the other two subsystems and activate information from long-term memory, it is the most important component of working memory, and plays a critical role in problem solving (Baddeley, 1992). But its function had not been clearly explored until recently. Swanson and his colleagues (2008) investigated the *central executive* in mathematical problem solving with kids in primary schools. They found the growth of the *central executive* was related to increasing problem solving performance.

The later development of the working memory model suggested that the *central executive* has no information storage capacity and there is another component, the episodic buffer, which provides temporary information storage for the *central executive* (Baddeley, 2000). This is the place holding information from both working memory and long-term memory, and also "binding information from a number of sources into coherent episodes" (pp. 421).

Recently, Cowan's Embedded Processes Theory of working memory is becoming popular (Cowan, 1999; 2005). In his theory, working memory is considered as activated long-term memory. The focus of attention with limited capacity activates information from long-term memory. The capacity of this attentional focus is the capacity of the working memory in his theory. This is a very different perspective on working memory. But in Baddeley's opinion (2012), the differences are mostly in the terminology and the areas of research focus.

Figure 4-3 The model of working memory (Baddeley, 2002)
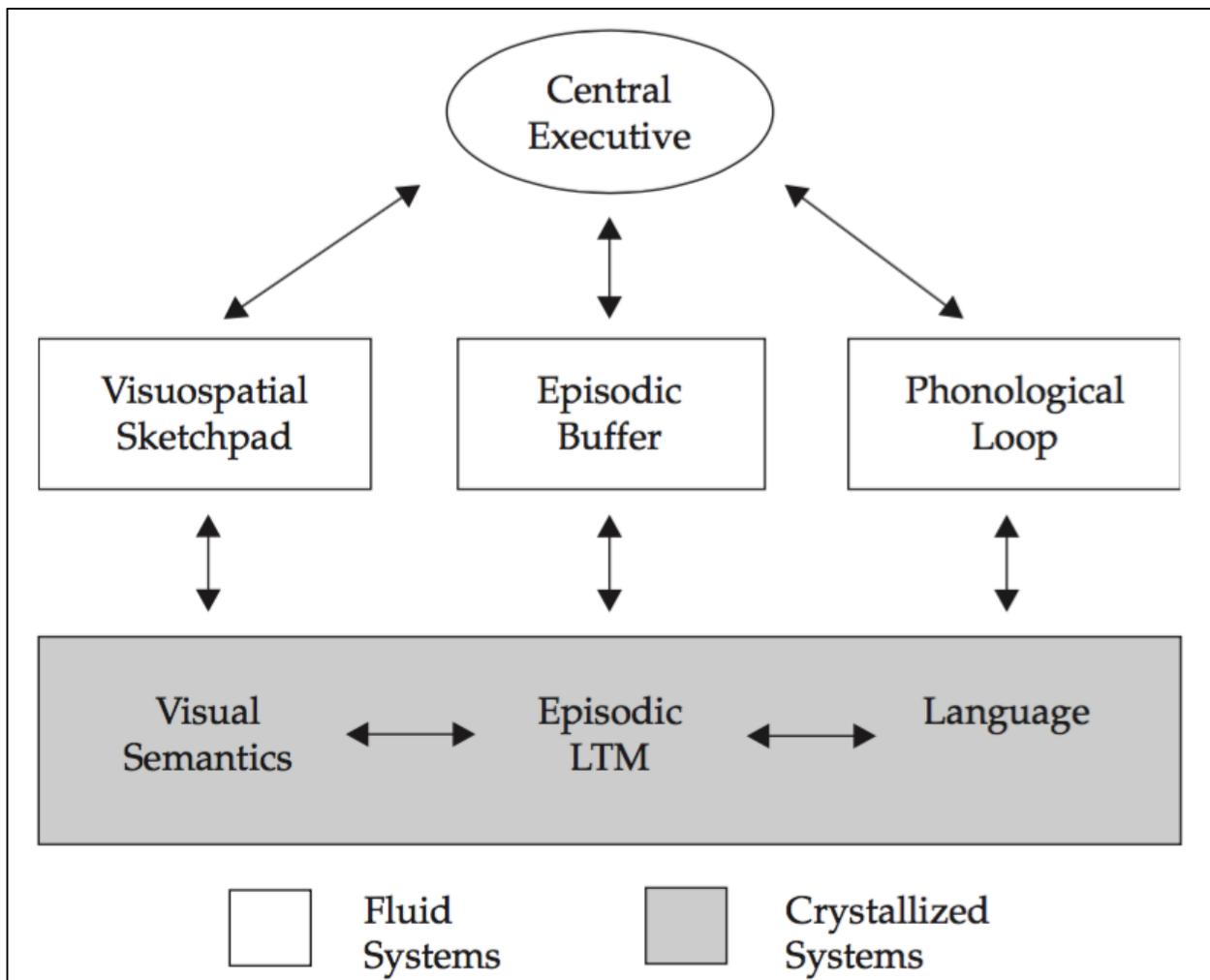
Models of Student Thinking in PER

The model we try to build is to qualitatively explain students' physics problem solving. It cannot be done without the concepts from physics education research. Physics education

researchers have devoted great effort toward understanding students' physics knowledge structure in the mind. The misconceptions view was proposed in the early 1980s (Clement, 1982; McDermott, 1984). In the area of physics education, it has been found that many students have difficulties in understanding physics concepts. The main obstacle is the conflict between students' pre-conceptions, and the systematic physics concepts that they need to learn. This view focuses on the naïve prior knowledge students hold, and the instructional interventions which could change it. The advocates of the misconceptions view claim many understandings that students bring to the classroom are misconceptions, such as a harder push makes an object move faster. These misconceptions need to be corrected. Posner and colleagues (1982) suggested a four-step intervention to change a misconception: (1) Generating a scenario to show the misconception does not work, (2) Introducing and elaborating the new concept, (3) Showing the new concept fits the scenario which has dissatisfied the misconception, and (4) Using the new concept in other scenarios.

The misconceptions view and the theory of concept change have many limitations. First, the term "misconception" entails the idea that students' prior knowledge is wrong. The responsibility of the instructor is to "correct" students' prior knowledge. It creates an unnecessary tension between students and instructors. Second, it does not provide a theory base to improve the effectiveness of instructional material. The misconceptions view describes students' difficulty in learning as "resistant to change". The underlying assumption is that students are unwilling to abandon their misconceptions and accept the physics concepts; therefore students are to be blamed when instruction fails.

Due to those limitations, many researchers looked for an alternative approach to understand students' prior knowledge. The model of *p-prims* suggested by diSessa (1993) was

very influential in this regard. Unlike the misconception view, which treats students' prior knowledge as a whole structure, the model of p-prim describes students' prior knowledge as a loose entity that is consisted by the basic particles, or *p-prims*. To novice students, *p-prims* are the small pieces of knowledge that cannot be further explained. Building upon the model of *p-prims*, Hammer proposed the resources theory (2000) to give a clear map of students' knowledge structure. In this theory, resources in a student's mind are similar to chunks of computer code. Resources by themselves are neither correct nor incorrect; it just depends on the context. For example, the idea of "closer is stronger" is a resource. It is correct when it is used to explain why people would feel hotter when getting closer to fire. But it is incorrect when it is used to explain why summer is hotter than winter.

The resources theory can also provide an explanation for misconceptions. Hammer argues that the so-called misconceptions are due to the nature of resources held by novice students. Students activate resources primarily based on their prior experience from everyday life, or the surface features of the context. However, both of these can be misleading, because many physics concepts and physics contexts are counter-intuitive.

Many teaching strategies have been generated based on the resources theory. Since the major issue facing students is activating inappropriate resources, the mediation is to create contexts that could promote the activation of productive resources. For example, the model generated by Podolefsky and Finkelstein (2007) adopted multiple interventions such as multimedia representation, analogy, and layering of meaning, in order to train their students to have a systematic understanding of electromagnetic waves, and they achieved positive results.

Significance of Study

This chapter proposed a model of physics problem solving with multimedia hints. Previous chapters showed that the current theory could not fully explain the questions that we are interested in. We tested a set of principles of multimedia learning (Mayer, 2005) and the multiple resources model (Wickens, 2002) in chapter 2, by analyzing the problem solving performance data. We found the data fit neither CTML, nor the multiple resources model completely. With the help of *cognitive load theory* (Sweller, 1988), we discussed the eye movement data in chapter 3, and found evidence to support the performance data, but there are still many questions without answers. The theoretical model proposed in this chapter will seek answers for those questions. The main purpose of this chapter is to provide theoretical explanations of the data presented in the previous chapters. Specifically, we address the following research questions:

1. How do theories provide explanations on the effects of presenting a single hint on problem solving performance?

2. How do theories provide explanations on the effects of presenting multi-modality hints on problem solving performance?

3. How do theories provide explanations on the effects of presenting multimedia hints on eye movements during problem solving?

The Model

The model describing the steps of problem solving with pictorial information and linguistic information is shown in Figure 4-4. It integrates the concepts of Mayer's (2002) CTML and Ohlsson's (1992) RCT.

Figure 4-4 A model of problem solving with the concepts from RCT (Ohlsson, 1992) and CTML (Mayer, 2002).
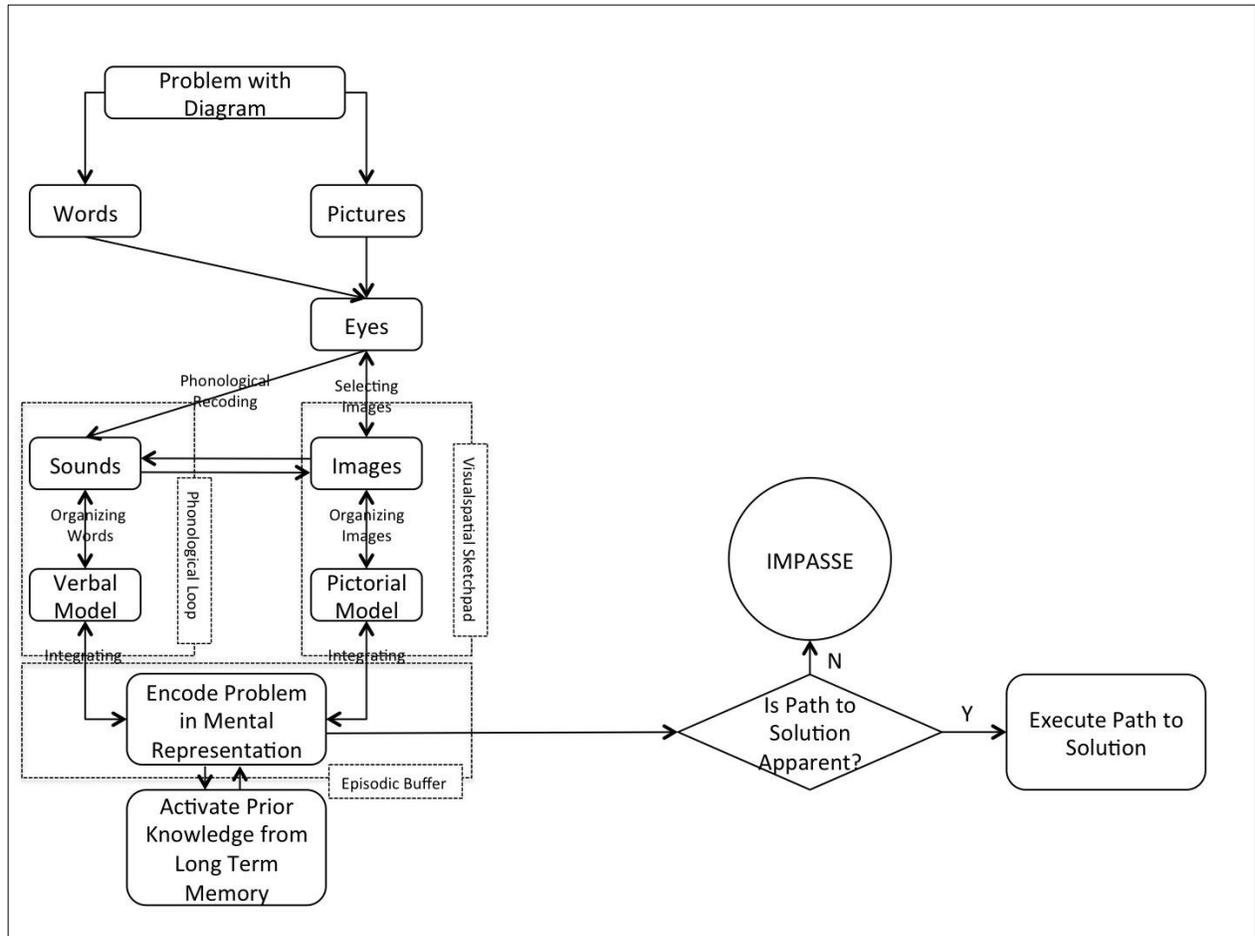
Since this model is for solving physics tasks presented on a computer screen, the information of the problem comes to the eyes as words and pictures. As a sensory system, eyes can hold the image of words and pictures for only a brief time period. Then the words go through the process of phonological recoding, and are stored in the *phonological loop* of working memory. Pictures are stored in the working memory's *visuospatial sketchpad*. Working memory actively processes the information and turns the raw information (i.e., sounds and images) into the constructed knowledge (i.e., verbal model and pictorial model). The arrow from sounds to images represents the reference from statement of the problem, to a diagram of the problem. For example, one of the ball problems in our study reads, "Two balls roll along the paths shown."

Participants may inspect the visual representation of "balls" and "paths" when they read these two words. Also, the arrow from images to sounds represents the reference from the diagram of the problem to the statement of the problem. Organizing and constructing the verbal and pictorial models are not two separate processes. They are two parallel processes affecting each other. As the result of knowledge construction, the mental representation of the problem is established. This is a critical step of problem solving. This mental representation determines prior knowledge activation from long-term memory and the path to solutions. It needs to be clarified that building the mental representation upon the external information, and the effect of the mental representation on prior-knowledge activation, are not one-way. Altering the salience of the presented material, such as highlighting some words or some parts of the diagram, could affect how raw information is processed and how knowledge is constructed. On the other hand, the half-baked mental representation could also bias the processes of selecting and organizing information. This is also true for the relationship between mental representation and prior-knowledge activation. The mental representation influences which part of prior knowledge is activated, and the activated prior knowledge in turn can affect the type of mental representation established.

With the established mental representation, solvers look for a path to a solution. If the path to a solution is apparent, the solvers would construct and report this solution. If not, the solvers would be stuck with an impasse. According to Ohlsson's (1992) RCT, to break an impasse, students need to change the unproductive mental representation of the problem.

For the purposes of this study, the model presented in Figure 4-4 is expanded to include multimedia hints. The expanded model with hints is shown in Figure 4-5.

Figure 4-5 A model of problem solving with multimedia hints.
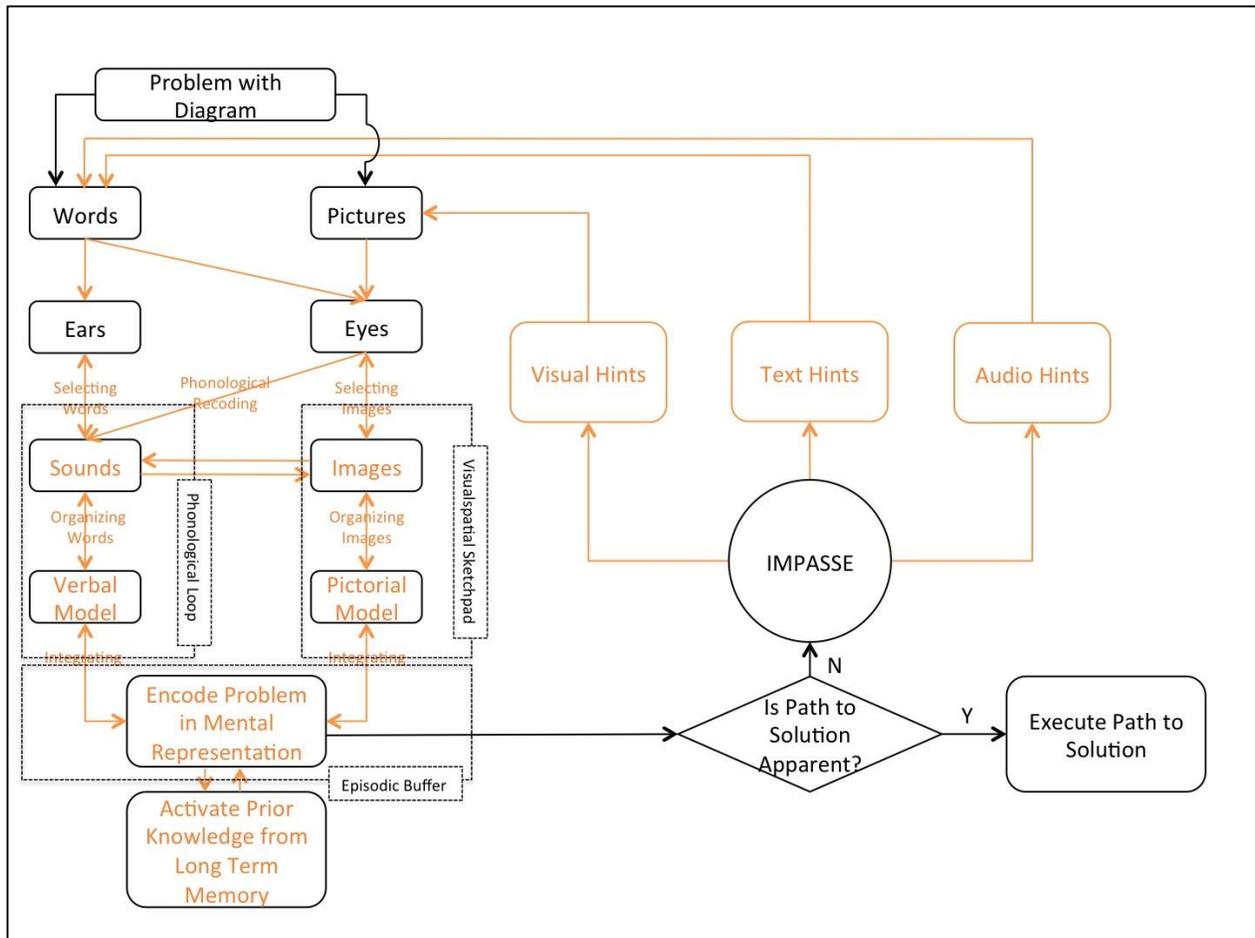
We include three types of hints as the intervention to break the impasse. They are categorized according to their presenting modalities and they are processed through different information processing channels (see Figure 4-5). Visual hints go to the eyes as pictures. They go through the right channel of the model to be processed as pictorial information. Audio hints go to the ears as sounds. They go through the left channel of the model to be processed as verbal information. Text hints are verbal information with visual representation. They are perceived by the eyes, and then transferred into sounds. Processing the information from text hints requires resources from both channels.

After they have reached the working memory, the functions of visual hints, text hints, and audio hints depend on their contents rather than modalities. As reviewed in the previous section

around the framework of attention cueing, they could help with selecting, organizing, and integrating information. As the result of affecting these three steps of information processing, the mental representation could be altered based on revised verbal, and pictorial models. Then this new mental representation could lead to a path to the correct solution.

Analysis and Discussion

A Summary of Results of Data Analysis

In the study of problem solving performance, we found text hints were better than audio hints. Visual hints were the type of hints that could significantly improve all other types of hints and their possible combinations. There was no significant difference between text + audio hints and text hints. There was also no significant difference between text + audio hints and audio hints. But the performance of text condition was marginally better than the performance of text + audio condition. And the performance of text + audio condition was marginally better than the performance of audio condition.

As for the study of eye movements, two groups of data were analyzed. The domain relative ratios were analyzed to explore how long participants inspected hints, and the mean fixation durations were analyzed to measure the cognitive load imposed by hints. We found a split attention effect when presenting visual hints and text hints simultaneously: participants attended to visual hints for longer time than they attended to text hints. This preference was not affected by adding audio hints. A further analysis showed that hint presentation time was long enough for participants to finish both reading text hints, and inspecting visual hints. As for the analysis of mean fixation durations, we found text hints imposed the least cognitive load when they were presented, but the cognitive load after presenting text hints had an increasing trend. Audio hints imposed higher cognitive load when they were presented and the cognitive load

remained high afterward. The cognitive load imposed by visual hints cannot be directly measured by the mean fixation duration. The comparison of mean fixation durations between visual condition and text condition cannot directly reflect the comparison of cognitive load between these two conditions. Since the visualization presented by visual hints and the on-screen words presented by text hints were two different visual stimuli. But comparing the mean fixation durations during hints and after hints for visual hints suggested there was a clear decreasing trend on cognitive load from during hints to after hints.

The Effect of Presenting A Single Hint on Problem Solving Performance

Visual hints are processed by the right channel of the system as shown in Figure 4-6.

Figure 4-6 A model of problem solving with visual hints
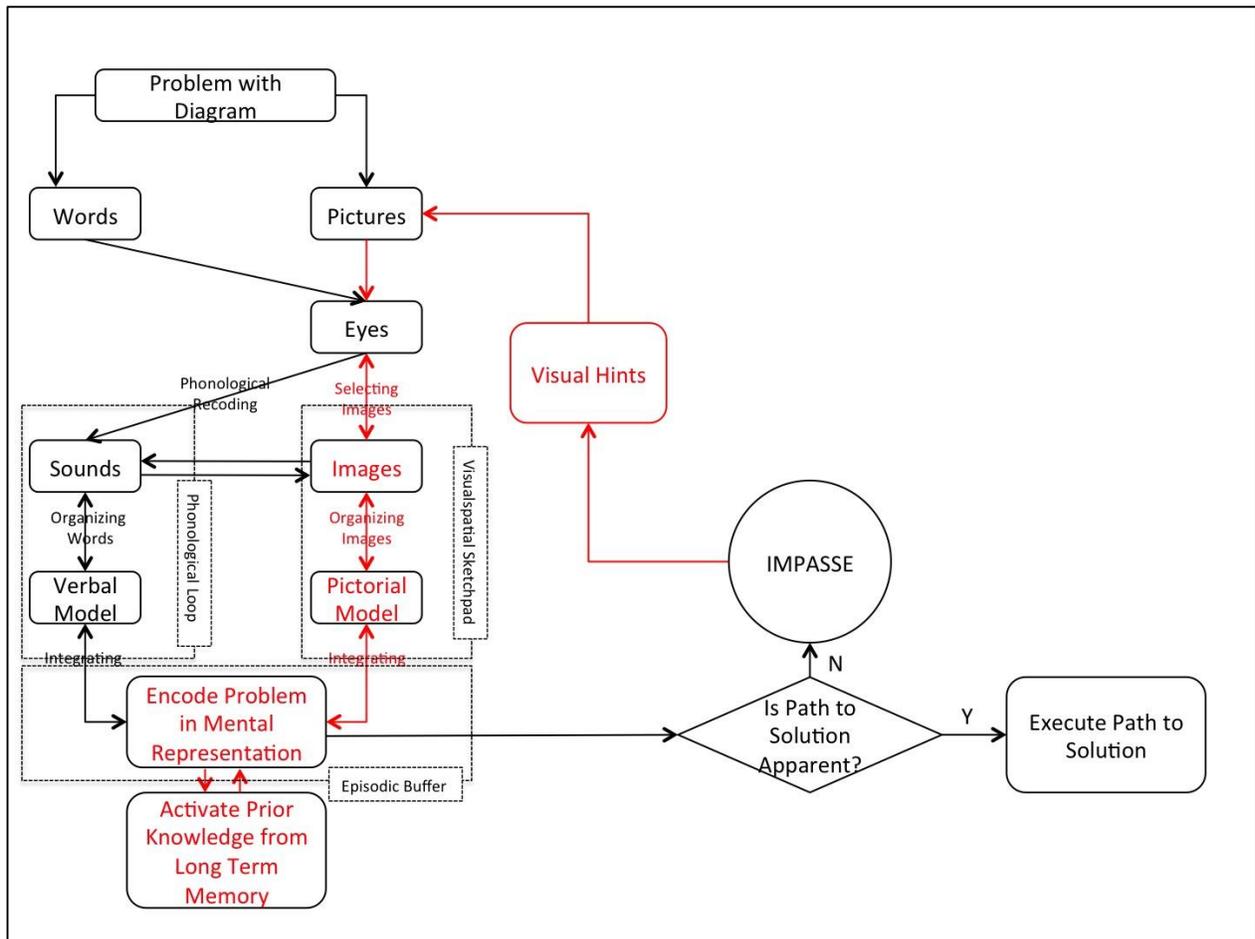
Visual hints in our study highlighted the correct solution areas of pictures in a flashing manner. The salience of visual hints drew participants' visual attention toward these highlighted areas. According to Hammer's resources theory, participants without hints would be attracted by the surface features of tasks or the features aligning with their prior everyday experience. But with the help of visual hints, participants are more likely to select the pictorial information from the correct solution areas. Visual hints also helped to organize and integrate pictorial information in order to build a new pictorial model, and also assisted in the construction of a new mental representation. For example, the visual hint on the ball tasks highlighted the distances between two adjacent balls. This trigged participants to compare these distances, which in turn changed

106

their mental representation from "looking for positions" to "looking for distances", which is more likely to produce the correct solution.

As linguistic hints, the functions of text hints (see Figure 4-7) and audio hints (see Figure 4-8) are different from visual hints.

Figure 4-7 A model of problem solving with text hints
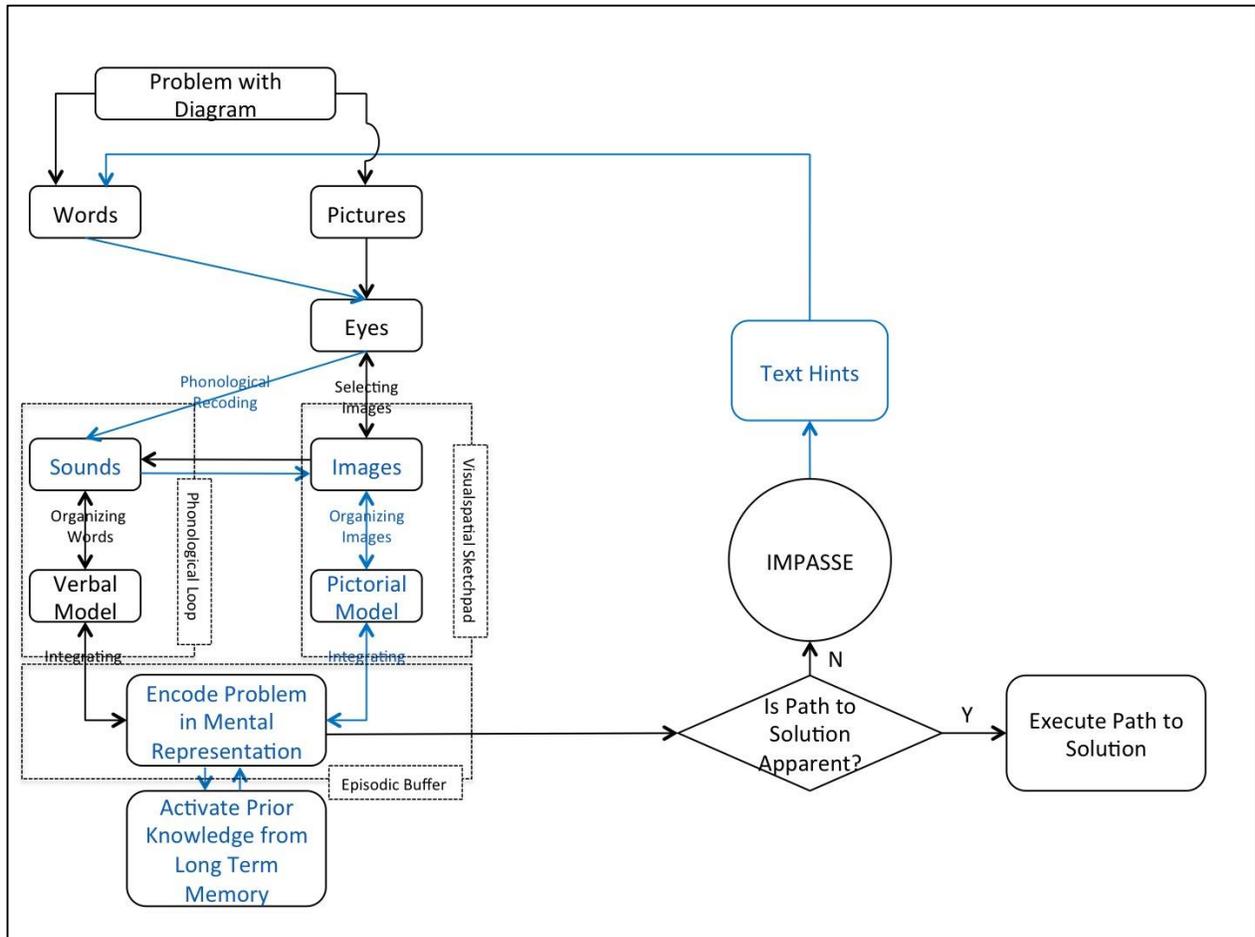
Figure 4-8 A model of problem solving with audio hints
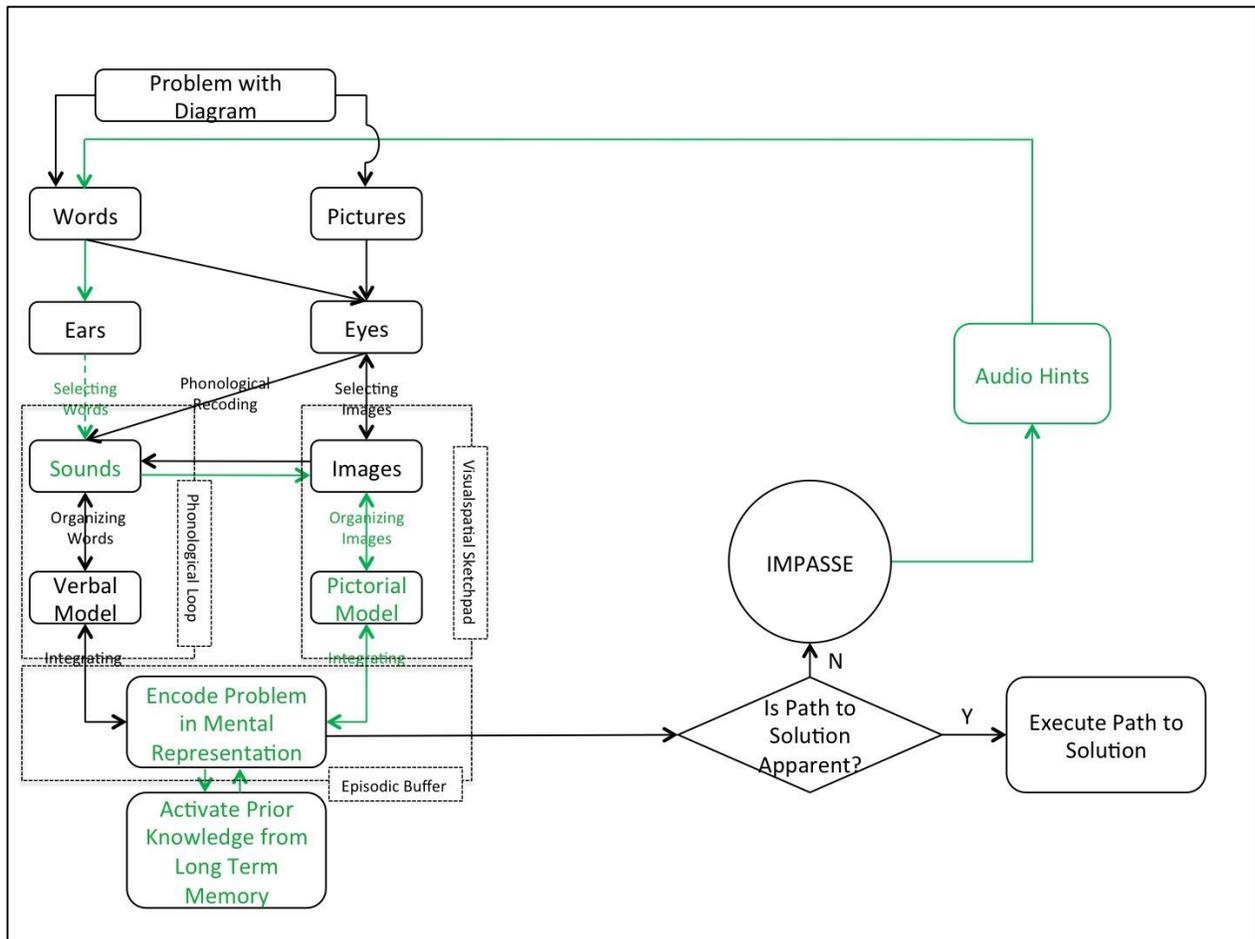
The commonalities of text hints and audio hints in our study are to be discussed first. Essentially, these two types of hints mimicked the content and the functions of visual hints. When these three types of hints were designed in our study, we started from visual hints. Text hints and audio hints were generated to match the content of the visual hints. As discussed previously, the functions of our visual hints are; selecting, organizing, and integrating pictorial information. These three functions should be expected as the functions of text hints and audio hints. In our study, text hints and audio hints did offer help in organizing and integrating pictorial information to participants. For example, the content of text and audio hints for the graph tasks was "compare the slope along the curve at each moment in time". It explicitly told participants

the action they should take to organize and integrate the visual elements of the diagrams. But the function of selection is questionable. Like the hints for graph tasks we listed above, the words "slope" and "curve" implicitly told participants the visual elements to select from the diagram, but for participants without proper physics knowledge, these two terms did not show any specific place of the diagram to look at. Moreover, the studies of linguistic selecting hints we reviewed before were all using direction words to guide participants' visual attention (Hommel, *et al.*, 2001; Ho and Spence, 2005). There were no direction words in our text and audio hints. Therefore, there is no evidence showing these two types of hints in our study could help in selecting pictorial information as visual hints did. They also cannot help with selecting, organizing, and integrating verbal information since neither of them are the purpose of visual hints.

The difference between text and audio hints is the modality. Text hints are on-screen sentences perceived by the eyes. They are images of words in sensory memory. Then the images of words are recoded from orthographic information to phonological information. Processing audio hints does not involve the step of recoding. Ears directly select the sounds of words. This demonstrates that there is no difference between processing text hints and processing audio hints beyond sensory memory.

Moreover, this model could also explain why text hints were better than audio hints. Text hints were stored in the memory of the eyes, then recoded as sounds. Audio hints were stored in the memory of the ears, then directly sent to the phonological subsystem of working memory. The path from the ears to working memory is transient as the sounds of audio hints were transient. For both text hints and visual hints, it is easy to select multiple elements or even re-select elements since all elements are equally accessible. But for audio hints, once one word

passed, there was no way to go back to hear that word once again. Participants have to either store this piece of information in *phonological loop*, which would cause extraneous cognitive load, or give up on it, which would compromise the hint benefits.

The Effect of Presenting Multi-modality Hints on Problem Solving Performance

The major benefit of presenting linguistic hints is that it could enhance the effects of visual hints on organization and integration. A good analogy here is adding two waves. The energy of each wave might be limited, but if these two waves are in phase, the resultant energy could be as much as four times the energy of each wave. The benefit of combining visual and linguistic hints is the reason that visual + audio hints were marginally better than visual hints, and visual + text hints were significantly better than just visual hints alone (see Chapter 2). The advantage of text hints over audio hints was still due to the limitation of selection words from the temporary sounds.

Figure 4-9 presents visual and text hints simultaneously, and Figure 4-10 presents visual and audio hints simultaneously.

Figure 4-9 A model of problem solving with visual + text hints.
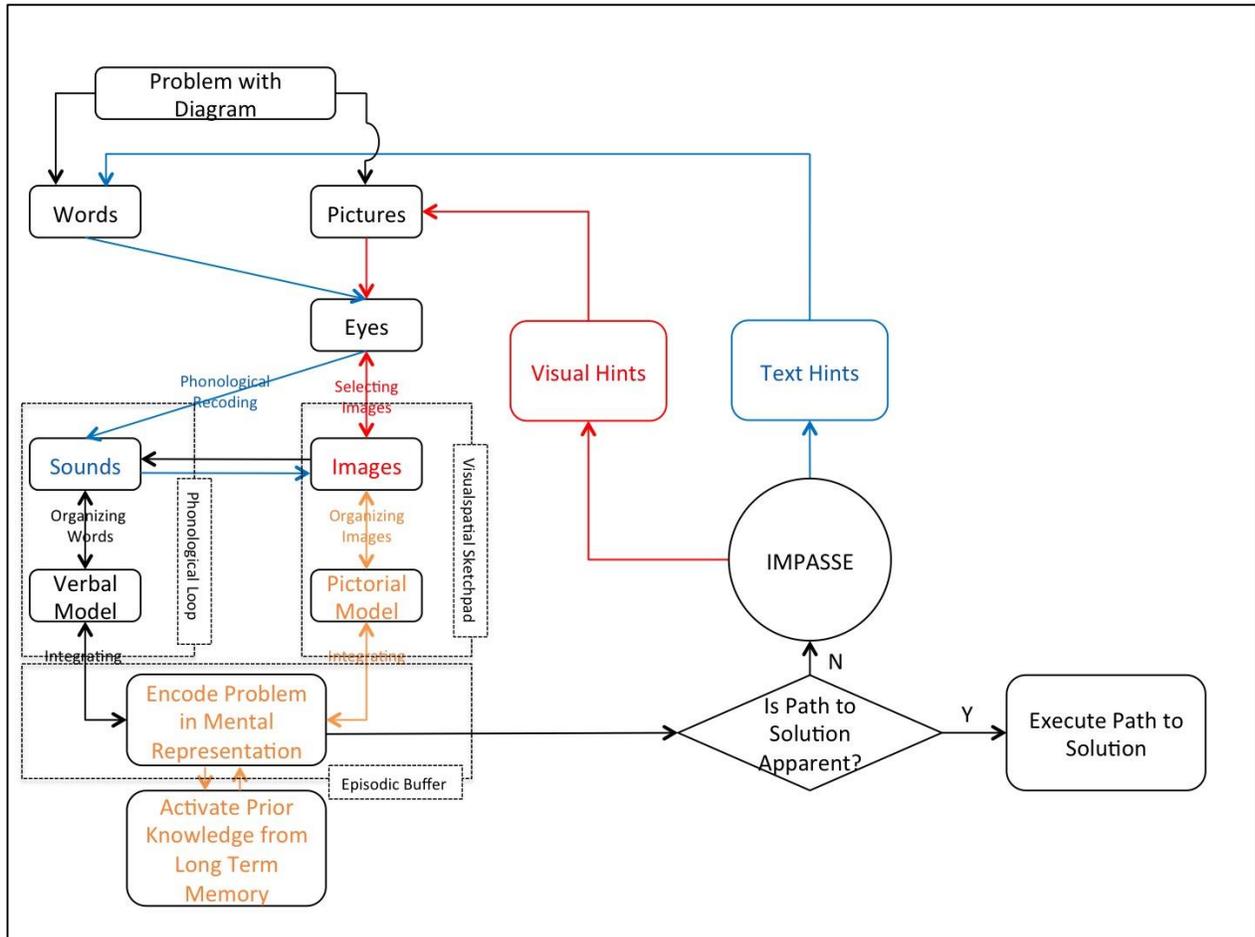
Figure 4-10 A model of problem solving with visual + audio hints.
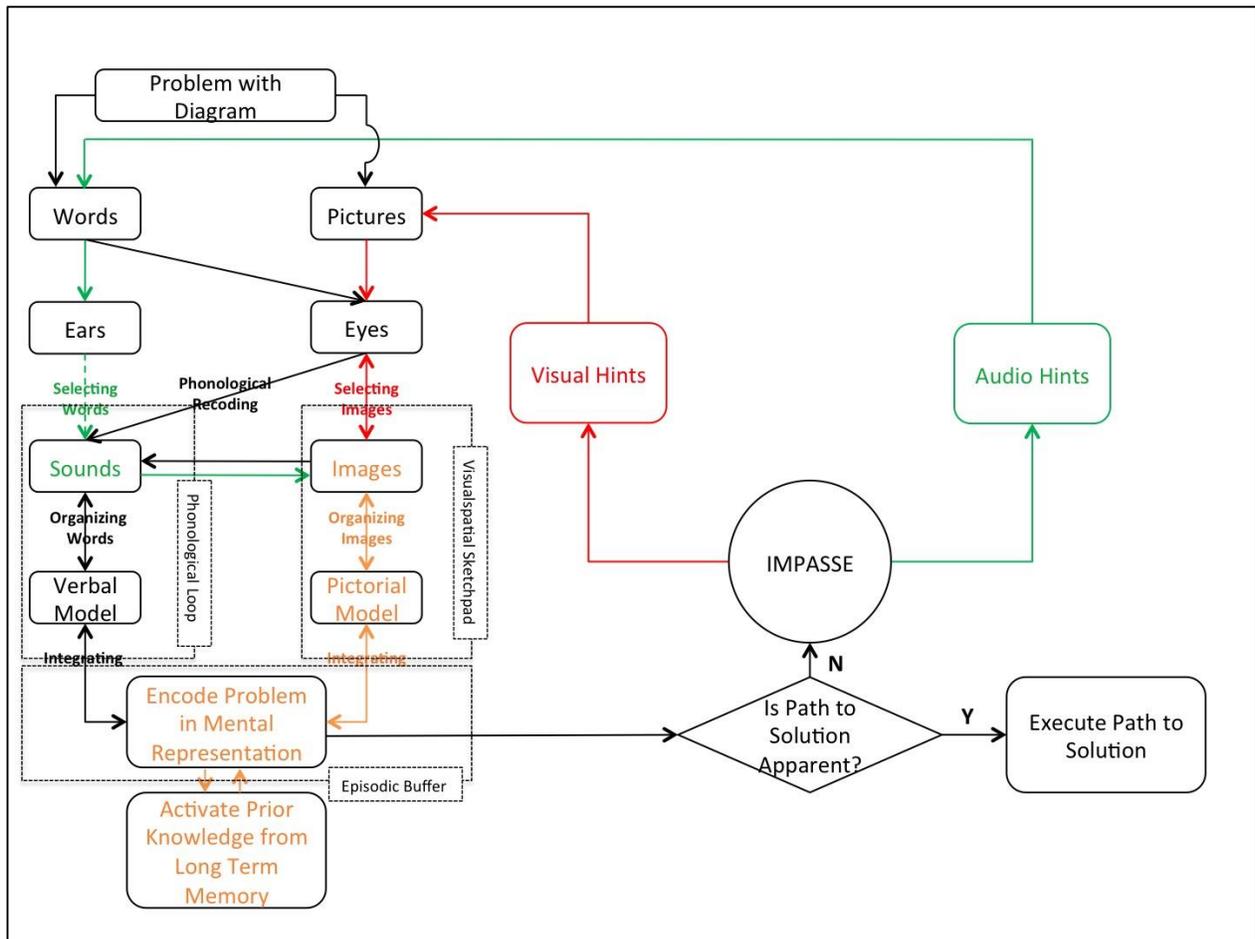
The fact that both visual hints and text hints are perceived by the eyes may cause an overload on sensory memory. This negative effect could be easily resolved by expanding the presentation time, or slowing the pace of the hints. One may argue that the negative effect of transient audio hints could also be resolved by prompting students to play audio hints multiple times, but due to the fact that we found that students were extremely unwilling to view hints multiple times, we still prefer text hints to audio hints.

Figure 4-10 presents two linguistic hints – audio and text -- simultaneously.

Figure 4-11 A model of problem solving with text + audio hints
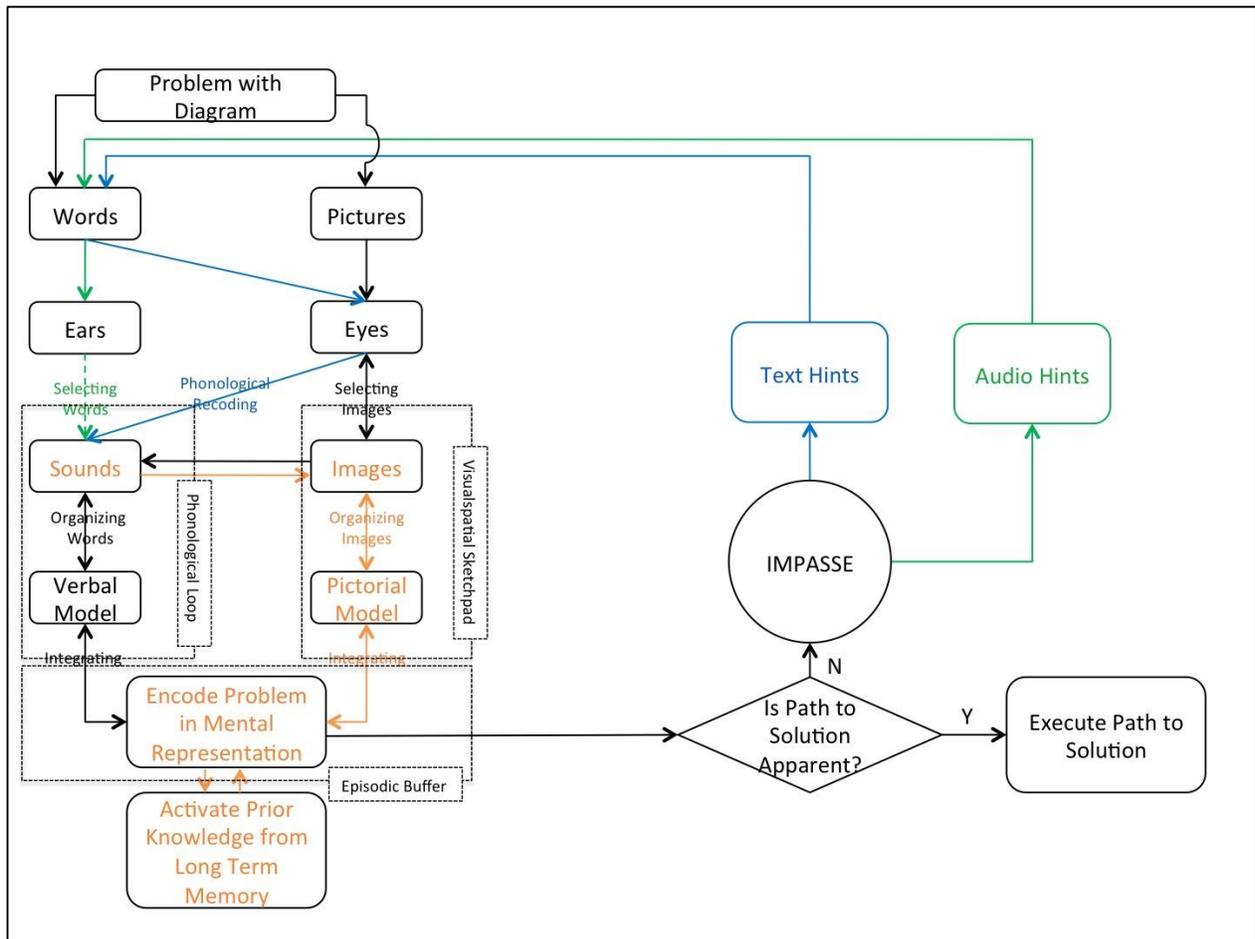
As we can see, the fact that linguistic hints cannot directly select pictorial information for participants, is not eliminated by adding text and audio hints together. Presenting text hints with audio hints does benefit audio hints because the information is not transient any more, but presenting audio hints with text hints does nothing to assist text hints, since everything audio hints offer participants can easily obtain from text hints. Moreover, audio hints even weaken text hints by overloading the linguistic channel of information processing. This discussion explains the reasons that text + audio hints were slightly better than audio hints, and slightly worse than text hints.

The presentation of all three hints together is depicted in Figure 4-12.

Figure 4-12 A model of problem solving with visual + text + audio hints.
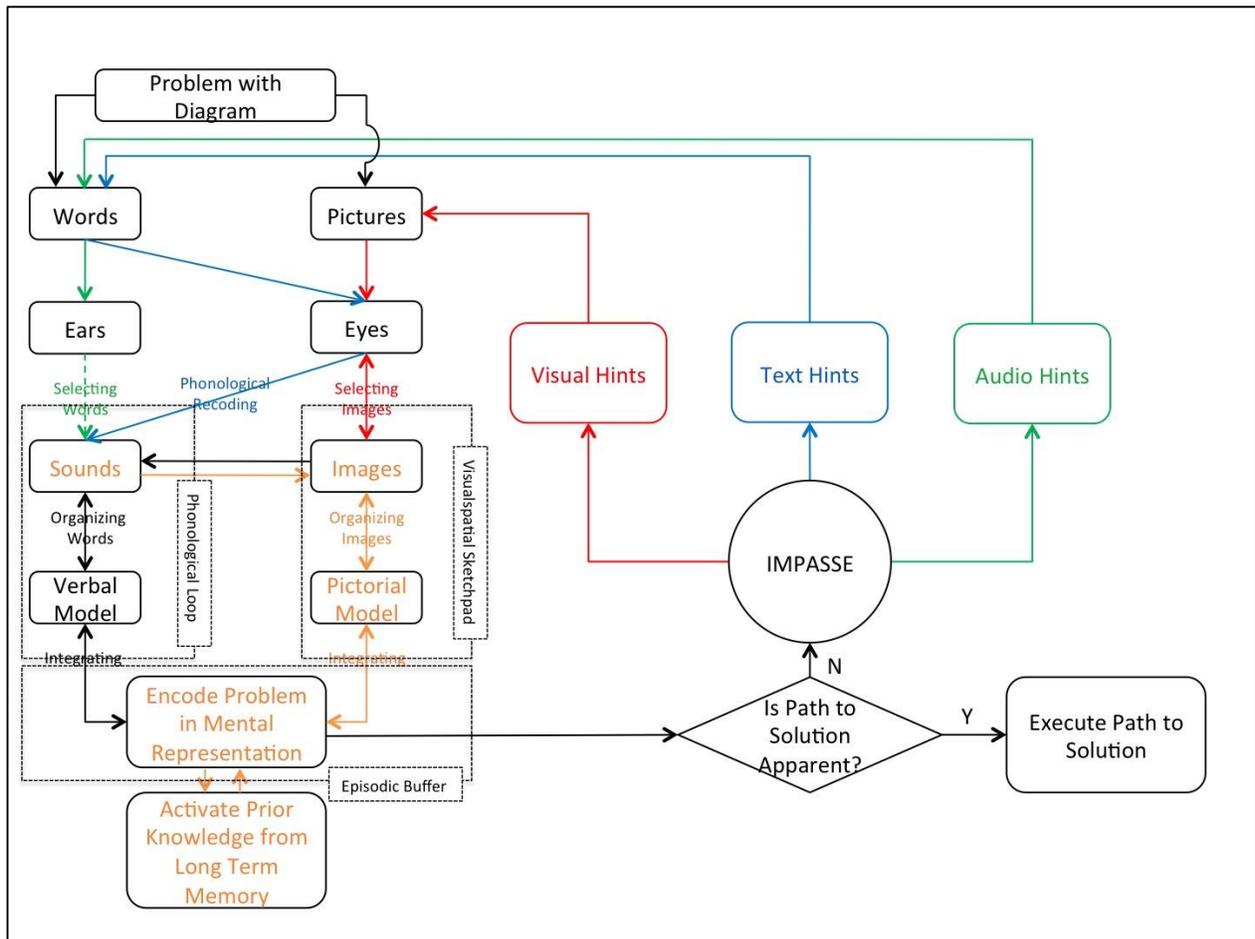
Based on our previous discussion, combining visual and text hints are the most effective combination so far. Adding audio hints would increase the cognitive load of processing linguistic information. Therefore, visual + text + audio hints should underperform visual + text hints, and our performance data aligns with this prediction.

The Effect of Presenting Multimedia Hints on Eye Movements

The previous two sections have discussed how hints affect problem-solving performance. This section will focus on the effects of hints on visual attention.

The problem-solving model we proposed has discussed the advantage of visual hints. Visual hints could help select correct solution relevant visual elements directly from diagrams.

This was supported by the eye movement data reported in Chapter 3. First, we found participants preferred visual hints to text hints when these two hints were presented simultaneously. The domain relative ratio on visual hints was significantly larger than the domain relative ratio on text hints, and there was a clear decreasing trend on the domain relation ratios on text hints from the first training tasks to the last training tasks. But there was no such trend for visual hints. All these effects were not affected by audio hints. These evidences suggested that participants did follow visual hints closely across the whole training process, and they attended to the visual elements highlighted. They were not so loyal to text hints. The explanation we proposed is that the functions of organizing and integrating text hints are identical for all isomorphic training tasks. Participants did not need to attend to text hints once they felt they knew what to do. But the function of selecting offered only by visual hints was unique for each training task since the features of diagrams were always changing. The participants had to attend to visual hints every time to obtain the assistance of selection from visual hints.

Second, by measuring mean fixation durations, we found the participants with visual hints had a significant increasing trend on cognitive load from the time interval before hints to the time interval during hints, and a significant decreasing trend on cognitive load from the time interval during hints, to the time interval after hints, since visual hints distinguished the correct solution relevant elements from the whole picture. They forced the participants with visual hints to facilitate and complete the pictorial information selection during the presentation of visual hints. This would increase the cognitive load from before hints to during hints, and also decrease the cognitive load from during hints to after hints. After visual hints were presented, the participants just needed to organize and integrate information to construct the mental representation. But for the participants without visual hints, they had to engage with selection,

organization and integration through the whole problem solving procedure until reporting

solutions. Therefore, the conditions without visual hints did not have such trends on cognitive

load change.

As for text hints and audio hints, the model suggested that the information from audio

hints had to be stored in the *phonological loop* of working memory for real-time accessing due to

the audio modality's transient nature. On the other hand, text hints did not need to be stored in

the *visuospatial sketchpad* of working memory since each term of text hints was equally

presented and accessible. This discussion was confirmed by the comparison of the cognitive load

between the text, and audio conditions (see Chapter 3). Audio hints did impose higher cognitive

load than text hints. This suggested that the participants with audio hints did exert more effort

than the participants with text hints.

Limitations of the Model

This model is a theoretical effort to explain the problem solving performance and eye

movement data in our experiment. This model serves more like a starting point to the

development of a theoretical understanding of conceptual physics problem solving with hints. To

validate our further discussion, the boundaries and limitations of this model should be well

defined.

First, this model covers conceptual physics problem solving with pictorial information

specifically. The group of tasks we used in this study and some of our previous studies all have

two features: (1) numerical calculation is not needed, and (2) solvers have to draw the

information from diagrams to successfully solve the task. Solving the physics tasks with heavily-

loaded calculations is not discussed in this model, since using external devices (i.e., pencil and

paper, calculator, or computer) to record the problem solving steps clearly violates the basic idea of this model.

Second, this model discusses individual problem solving only. There are many studies analyzing group problem-solving scenarios in the PER community (Heller, *et al.*, 1992; Ploetzner, *et al.*, 1999; Harskamp and Ding, 2006). The interpersonal collaboration aspect on problem solving is clearly not within the scope of this model. Here, we only deal with the case which one solver is facing one physics task each time.

Third, this model currently focuses on primarily explaining the interaction between physics problem solving, and multimedia hints. According to the literature review above, there are several key components in a problem-solving model: mental representation, impasse, insight, and solution. The data we collected in this study provides evidence only for the mechanism that deals with how hints were processed. This model cannot differentiate between types of mental representation, impasse, and insight. This model also cannot differentiate mechanisms for breaking impasse.

Chapter 5 - Conclusions

Overview of Research

Physics problem solving is one of the critical subfields of PER (Hsu, *et al*., 2004; Docktor and Mestre, 2014). Understanding how students solve problems and searching for innovative instructional strategies to help students learn problem solving has always motivated physics education researchers. Due to newly emerging technologies, both research and teaching methods have gone through a great revolution. As educators, we embrace the convenience of computer-aided instruction. As researchers and developers, we would like to know how students learn using computer-aided instruction, and how to better design computer-based instructional material to help students learn physics problem solving. This research was inspired by many studies in cognitive psychology that tried to understand human problem solving. We found tracking students' eye movements is a plausible approach to explore their cognitive activities during problem solving (Knoblich, *et al*., 2001; Grant and Spivey, 2003).

Using eye movement data to study how participants solve conceptual physics tasks has been our group's research focus since the first study done by Madsen *et al*. (2013). We found that correct solvers and incorrect solvers looked at different areas of the diagram on some tasks. Then in a follow up study (Rouinfar, 2014), we found using visual cues to train participants to look at the areas of a diagram, where the correct solvers attend, could improve participants' problem solving performance. In the current study, we investigated the influence of multimedia hints on problem solving performance and visual attention. We examined students' performance and eye movements on solving conceptual physics tasks with pictorial representations. A theoretical model was built to conceptually understand students' problem-solving procedures and the effects of multimedia hints on both problem solving performance and visual attention.

Overview of Theory and Method

We mainly adopted Ohlsson's Representational Change Theory (RCT, Ohlsson, 2002) as the model to depict cognitive activities in problem solving. The framework of attention cueing (de Koning, *et al*., 2009) was used to understand the function of hints. The representation of hints was designed according to the design principles of Cognitive Theory of Multimedia Learning (CTML)(Mayer, 2002, 2005b). We used a full factorial design: 2 (visual hint / No visual hint) × 2 (text hint / No text hint) × 2 (audio hint / No audio hint). Participants were randomly assigned to one of the eight conditions. CTML and Wickens' (2002) Multiple Resources Theory (MRT) are the two major theories we tested to gauge participants' problem solving performance. As for eye movement data, we tested hypotheses generated on Mayer's CTML, Wickens' MRT, and empirical evidences (d'Ydewalle *et al*., 1991; Latorella, 1998; Schmidt-Weigand, *et al.,* 2010b; Kalyuga, *et al*., 2001). Next, to provide a coherent explanation to both problem solving performance and eye movements, we combined Ohlsson's RCT (2002), Mayer's CTML (2002) and the framework of attention cueing (de Koning *et al*., 2009) as the main body of our problem solving model with multimedia hints. This model also has benefited from amalgamating aspects of cognitive load theory (Sweller, 1988), the model of working memory (Baddeley, 1992); and resources theory (Hammer, 2000) which is commonly used in PER.

Addressing Research Questions

Research Question 1

In Chapter 2 we investigated the problem solving performance of students in each of the eight conditions. The first research question asked, "How does the combination of visual, text, and audio hints affect students' performance on solving introductory conceptual physics tasks with graphic representation?" We found that working through training tasks with hints could

help students improve their problem solving performance, but there were no significant differences between conditions on the performance of near and far transfer tasks. There were significant differences between conditions on training tasks' performance. Participants from the visual + text condition outperformed those from the text condition, and participants from the visual + text + audio condition outperformed those from the text + audio condition. These two evidences were consistent with the hypotheses generated by the *multimedia principle* of Mayer's CTML. The competing hypotheses generated by Wickens' (2002) MRT, were not supported by the evidence. The visual condition outperformed the no hint condition. This was consistent with the hypothesis generated by the *signaling principle* of Mayer's CTML. The text condition significantly outperformed the audio condition on the training tasks. This contradicted the hypothesis generated by the *modality principle*. In other words, we found evidence to support the *reverse modality effect*. We also found that the visual + audio condition outperformed the visual condition and the audio condition. This was consistent with the hypothesis supported by Wickens' MRT. Post hoc tests compared the problem solving performance on training tasks for those who solved initial tasks incorrectly from the visual, visual + text, and visual + audio conditions. We found the visual + text condition significantly outperformed the visual condition for this group of students, and there was no significant difference between the visual + audio condition and the visual condition. This was contrary to the *modality principle* of Mayer's CTML. In summary, we found neither Mayer's CTML nor Wickens' MRT can provide coherent explanations for the data of problem solving performance in our study.

Research Question 2

There were two sub-questions for the second research question. The first one asked how students split their visual attention on attending to multi-modality hints, and the second one was

about students' cognitive load in problem solving. In Chapter 3 we investigated the effect of multimedia hints on students' visual attention. For the first sub-question, we found students split their visual attention when visual hints and text hints were presented simultaneously. This split-attention effect was not affected by audio hints. A post hoc test on the numbers of fixations per content word of text hints was conducted. We found the students in our study had a sufficient amount of time to read the text hints, and inspect the visual hints. Adding text hints to visual hints provided one more resource of help information. This evidence partially explained the reason the visual + text condition outperformed the visual condition.

As for the second sub-question, we investigated students' cognitive load by studying their eyes' fixation durations. We found text hints imposed a significantly less cognitive load than audio hints. This evidence revealed the reason why text hints were better than audio hints. We also explored the relationship between hint design and cognitive load imposed by hints.

A post hoc discussion compared the cognitive load before hints, during hints, and after hints for each of the eight conditions. We found visual hints were the hints with a clear upward trend from before hints to during hints; and a clear downward trend from during hints to after hints. This evidence was explained in answering the third research question.

Research Question 3

The third research question was discussed in Chapter 4. The goal was to construct a conceptual model using existing theories such as Mayer's CTML (2002), Ohlsson's RCT (1992) and Wickens' MRT (2002) in the context of the three cognitive processes related to hints: selection, organization, and integration. There were three sub-questions for this research question. The first sub-question asked how these theories provide explanations for the effects of presenting a single hint on problem solving performance. We discussed three hint functions:

121

selection, organization, and integration for each of the visual hints, text hints, and audio hints. In comparison to text hints and audio hints, the advantage of visual hints is it could directly help participants select visual elements that are relevant to correct solutions. Text hints are a better resource of help information than audio hints due to audio hints' transient nature. The second sub-question asked how these theories provide explanations for the effects of presenting multi-modality hints on problem solving performance. The fact that presenting visual hints and text hints together could enhance the functions of organization and integration via providing both visual and linguistic information, makes the visual + text hint combination the most effective hint combination for improving conceptual problem solving performance. The third sub-question asked how these theories provide explanations for the effects of presenting multimedia hints on eye movements during problem solving. The function of selection exclusively offered by visual hints is customized on the visual features of each diagram. The functions of organization and integration offered by text hints are the same across all isomorphic training tasks within each task set. This is the reason participants always prefer visual hints, and the heavy cognitive load during visual hint presentation could be explained by visual hints forcing participants to complete the selection with hints.

In summary, the model we proposed provided coherent explanations to the effect of presenting visual hints, text hints, and audio hints separately and together. The model explained both the problem solving performance and eye movements of this study. The limitations of the model were also discussed.

Limitations and Future Work

One limitation of this study is that there is no performance difference on near transfer problems or far transfer problems between all eight conditions. These results are not consistent

with the results of our previous work (Rouinfar, 2014). We believe that our contradictory results might relate to differences in participants' physics background. The participants in Rouinfar's study were recruited from an algebra-based physics course for life science majors. Most of these students had completed high school physics. However, the participants we recruited in this study were from a conceptual physics course. Very few of them have had any prior physics course in high school, and none had a college physics course. Therefore, they lacked problem-solving experience and may not have seen similarities between transfer problems and training problems as related since these problems are different in representation and content. In future work, we need to have clearly articulated criteria for design of the near and far transfer problems. These problems need to be more carefully redesigned to probe learners' transfer performance. Furthermore, we may need to consider repeating our study with different audiences, perhaps ones that have physics background similar to those in our previous study (Rouinfar, 2014).

Another approach to address this issue is to provide feedback (i.e., telling students whether their answer is correct or incorrect right after they report it). According to our group's previous studies (Rouinfar, 2014; Agra, 2015), providing feedback to participants may improve students' problem solving performance. This research approach could also help us probe the interaction between hint modality and feedback.

Moreover, investigating eye movement data suggested that the sufficient amount of time we provided for participants to access hints was the reason that the visual + text hints condition helped problem solving better than the visual hints condition, since there was one more information resource to attend to. The follow up study could compare a short presentation time versus a long presentation time. The short presentation time should be set short enough that participants can barely finish reading the text hints. This could force participants to attend to

information via one hint modality only. With this research approach, we could explore which hint modality students prefer under the time constraint.

Anticipated Boarder Impacts

The work on multimedia hints described above may shed light on designing instructional intervention on helping student solve physics problems. Visual hints should be the first choice of help information giving to students. Functions of selection, organization, and integration should be considered during hint design. For example, a visual hint to help students to draw a free body diagram could be providing arrows indicating directions of forces. This is an example of using visual hints to help students select specific visual elements (i.e., directions corresponding to forces) from the diagram. Much more work is necessary to better understand the cognitive underpinnings of multimedia hints and to apply this knowledge to the development of instructional materials and strategies.

References

Adams, W. K., & Wieman, C. E. (2015). Analyzing the many skills involved in solving complex physics problems. *American Journal of Physics*, *83*(5), 459-467.

Agra, E., Ph.D. thesis, Kansas State University, 2015.

Aldalalah, O. M., & Fong, S. F. (2010). Effects of modality and redundancy principles on the learning and attitude of a computer-based music theory lesson among Jordanian primary pupils. *International Education Studies*, *3*(3), p52.

Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction*, *19*(5), 376-386.

American Association for the Advancement of Science (2011). Vision and Change in Undergraduate Biology Education: A Call to Action, Washington, DC.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. The psychology of learning and motivation, 2, 89-195.

Baddeley, A. (1992). Working memory. Science, 255(5044), 556-559.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? Trends in cognitive sciences, 4(11), 417-423.

Baddeley, A. D. (2002). Is working memory still working?. *European psychologist*, *7*(2), 85.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, *4*(10), 829-839.

Baddeley, A.D. and Hitch, G.J. (1974) Working memory. In The Psychology of Learning and Motivation (Bower, G.A., ed.), pp. 47–89, Academic Press

Bahrick, H. P., Noble, M., & Fitts, P. M. (1954). Extra-task performance as a measure of

learning a primary task. *Journal of Experimental Psychology*,*48*(4), 298.

Banbury, S., & Berry, D. C. (1998). Disruption of office-related tasks by speech and

office noise. *British Journal of Psychology*, *89*(3), 499-517.

Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension.

TESOL quarterly, *24*(4), 746-753.

Boucheix, J. M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing

cues and internal continuous cues in learning with complex animations. *Learning and

instruction*, *20*(2), 123-135.

Briggs, G. E., Peters, G. L., & Fisher, R. P. (1972). On the locus of the divided-attention

effects. *Perception & Psychophysics*, *11*(4), 315-320.

Bromage, B. K., & Mayer, R. E. (1981). Relationship between what is remembered and

creative problem-solving performance in science learning. *Journal of Educational

Psychology*, *73*(4), 451.

Burr, D. C., & Morrone, C. (1996). Temporal impulse response functions for luminance

and colour during saccades. *Vision Research, 36*(14), 2069-2078.

Burr, D. C., Morrone, M. C., & Ross, J. (1994, October 6). Selective suppression of the

magnocellular visual pathway during saccadic eye movements. *Nature, 371*(6497), 511-513.

Butcher, K. R. (2014). 7 The Multimedia Principle. *The Cambridge handbook of

multimedia learning*, 174.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction.

Cognition and instruction, 8(4), 293-332.

Chekaluk, E., & Llewellyn, K. R. (1994). Masking effects in saccadic eye movements. In G. d'Ydewalle & J. Van Rensbergen (Eds.), *Visual and oculomotor functions: Advances in eye movement research* (Vol. 5, pp. 45-54). Amsterdam, Netherlands: North-Holland/Elsevier Science Publishers.

Chi, M. T., & VanLehn, K. A. (1991). The content of physics self-explanations. *The Journal of the Learning Sciences*, *1*(1), 69-105.

Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices*. Cognitive science, 5(2), 121-152.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of physics*, *50*(1), 66-71.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic bulletin & review*, *12*(5), 769-786.

Corkill, A. J. (1992). Advance organizers: Facilitators of recall. *Educational Psychology Review*, *4*(1), 33-67.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. *Journal of Educational Psychology*, *94*(2), 428.

Cowan, N. (1999). An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*, *20*, 506.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*(1), 87-185.

Cowan, N. (2005). Working memory capacity. Hove, UK.

Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica*, *50*(3), 291-323.

Cuevas, H. M., Fiore, S. M., & Oser, R. L. (2002). Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science*, *30*(6), 433-464.

de Groot, A. D. (1978). *Thought and choice in chess* (Vol. 4). Walter de Gruyter.

de Jong, T., & Ferguson-Hessler, M. G. (1986). Cognitive structures of good and poor novice problem solvers in physics. Journal of educational psychology, 78(4), 279.

de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review*, *21*(2), 113-140.

de Koning, B. B., Tabbers, H. K., Rikers, R. M., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and instruction*, *20*(2), 111-122.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo–spatial memory. *Neuropsychologia*,*37*(10), 1189-1199.

Della Sala, S., & Logie, R. H. (2002). Neuropsychological impairments of visuo-spatial STM. *Handbook of memory disorders*.

Dhillon, A. S. (1998). Individual differences within problem‐solving strategies used in physics. *Science Education*, *82*(3), 379-405.

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and instruction*, *10*(2-3), 105-225.

Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics-Physics Education Research*, *10*(2), 020119.

Dow, G. T., & Mayer, R. E. (2004). Teaching students to solve insight problems: Evidence for domain specificity in creativity training. Creativity Research Journal, 16(4), 389-398.

Duncker, K., & Lees, L. S. (1945). On problem-solving. *Psychological monographs*, *58*(5), i.

d'Ydewalle, G., Praet, C., Verfaillie, K., & Van Rensbergen, J. (1991). Watching subtitled television automatic reading behavior. *Communication research*, *18*(5), 650-666.

Ericsson, K. A., & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*,*37*(4), 571-583.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*(4), 313-331.

Goldenfeld, N., & Kadanoff, L. P. (1999). Simple lessons from complexity. *Science*, *284*(5411), 87-89.

Goodman, M. J., Tijerina, L., Bents, F. D., & Wierwille, W. W. (1999). Using cellular telephones in vehicles: Safe or unsafe? *Transportation Human Factors*, *1*(1), 3-42.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE-Life Sciences Education*, *11*(4), 364-377.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving guiding attention guides thought. *Psychological Science*, *14*(5), 462-466.

Hammer, D. (2000). Student resources for learning introductory physics. American Journal of Physics, 68(S1), S52. doi:10.1119/1.19520

Hardiman, P. T., Dufresne, R., & Mestre, J. P. (1989). The relation between problem categorization and problem solving among experts and novices. Memory & Cognition, 17(5), 627-638.

Harskamp, E. G., Mayer, R. E., & Suhre, C. (2007). Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction*, *17*(5), 465-477.

Harskamp, E., & Ding, N. (2006). Structured collaboration versus individual learning in solving physics problems. *International journal of science education*, *28*(14), 1669-1688.

Heller, P., & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, *60*(7), 637-644.

Heller, P., Keith, R., & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, *60*(7), 627-636.

Ho, C., & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of experimental psychology: Applied*, *11*(3), 157.

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and instruction*, *17*(6), 722-738.

Hommel, B., Pratt, J., Colzato, L., & Godijn, R. (2001). Symbolic control of visual attention. *Psychological Science*, *12*(5), 360-365.

Hsu, L., Brewe, E., Foster, T. M., & Harper, K. A. (2004). Resource letter RPS-1: Research in problem solving. *American Journal of Physics*, *72*(9), 1147-1156.

Isreal, J. B., Wickens, C. D., & Donchin, E. (1979, October). The event-related brain potential as a selective index of display load. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 23, No. 1, pp. 558-562). Sage Publications.

Javal, E. (1879). Essai sur la physiologie de la lecture. Translated by Ciuffreda, K. J. & Bassil, N. (1990). Essay on the physiology of reading. Ophthalmology and Physiological Optometry, 10, 381-384

Jonassen, D. H. (2011). Learning to solve problems: A handbook for designing problem-solving learning environments. New York: Routledge.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review*, *87*(4), 329.

Kalyuga, S. (2012). Instructional benefits of spoken words: A review of cognitive load factors. *Educational Research Review*, *7*(2), 145-159.

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(1), 1-17.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied cognitive psychology*, *13*(4), 351-371.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational psychologist*, *41*(2), 75-86.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & cognition*, *29*(7), 1000-1009.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 554.

Latorella, K. A. (1998, October). Effects of modality on interrupted flight deck performance: Implications for data link. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 42, No. 1, pp. 87-91). SAGE Publications.

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, *25*(6), 943-951.

Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology*, *17*(4), 401-418.

Logie, R. H. (1995). Visuo-spatial working memory.

Logie, R. H. (2011). The functional organization and capacity limits of working memory. *Current Directions in Psychological Science*, *20*(4), 240-245.

Lorch, R. F., Lorch, E. P., & Inman, W. E. (1993). Effects of signaling topic structure on text recall. *Journal of Educational Psychology*, *85*(2), 281.

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. *The Cambridge handbook of multimedia learning*, *147*, 158.

Madsen, A. M., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2012). Differences in visual attention between those who correctly and incorrectly answer physics problems. Physical Review Special Topics - Physics Education Research, 8 (1), 010122, doi:10.1103/PhysRevSTPER.8.010122.

Madsen, A., Rouinfar, A., Larson, A. M., Loschky, L. C., & Rebello, N. S. (2013). Can short duration visual cues influence students' reasoning and eye movements in physics problems? Physical Review Special Topics-Physics Education Research, 9(2), 020104.

Mautone, P. D., & Mayer, R. E. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology*, *99*(3), 640.

Mayer, R. E. (2002). Multimedia learning. *Psychology of Learning and Motivation*, *41*, 85-139.

Mayer, R. E. (2005a). Introduction to multimedia learning. In Mayer, R. E. (Ed.), *The Cambridge handbook of multimedia learning* (pp. 1-16). Cambridge University Press.

Mayer, R. E. (2005b). Cognitive theory of multimedia learning. In Mayer, R. E. (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). Cambridge University Press.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and instruction*, *20*(2), 167-171.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of educational psychology*, *93*(2), 390.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of educational psychology*, *82*(4), 715.

Mayer, R. E., & Johnson, C. I. (2008). Revising the redundancy principle in multimedia learning. *Journal of Educational Psychology*, *100*(2), 380.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology*, *90*(2), 312.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, *38*(1), 43-52.

McDermott, L. C. (1984). Research on conceptual understanding in mechanics. *Physics Today*, *37*, 24-32.

Miller, G. A. (1956). The magical number seven, + or minus two: some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-Hall.

North, R. A., & Gopher, D. (1976). Measures of attention as predictors of flight performance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *18*(1), 1-14.

Ohlsson, S. (1984). Restructuring revisited. II. Scandinavian journal of psychology, 25(2), 117-129.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. Advances in the psychology of thinking, chap-1.

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in human behavior*, *26*(1), 110-117.

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology*, *84*(4), 429.

Paas, F. G., & Van Merriënboer, J. J. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational psychology review*, *6*(4), 351-371.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory.*Educational psychologist*, *38*(1), 63-71.

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010, March). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 symposium on eye-tracking research & applications* (pp. 141-144). ACM.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). Organizing Instruction and Study to Improve Student Learning. IES Practice Guide. NCER 2007-2004. *National Center for Education Research.*

Paus, T., Marrett, S., Worsley, K. J., & Evans, A. C. (1995). Extraretinal Modulation of Cerebral Blood Flow in the Human Visual Cortex - Implications For Saccadic Suppression. *Journal of Neurophysiology, 74*(5), 2179-2183.

Ploetzner, R., Dillenbourg, P., Preier, M., & Traum, D. (1999). Learning by explaining to oneself and to others. *Collaborative learning: Cognitive and computational approaches*, 103-121.

Podolefsky, N. S., & Finkelstein, N. D. (2007). Analogical scaffolding and the learning of abstract ideas in physics: Empirical studies. *Physical Review Special Topics-Physics Education Research*, *3*(2), 020104.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science education*, *66*(2), 211-227.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372.

Rayner, K., & McConkie, G. W. (1976). What guides a reader's eye movements? *Vision research*, *16*(8), 829-837.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, *7*(3), 219.

Redish, E. F. (2004). A theoretical framework for physics education research: Modeling student thinking. *arXiv preprint physics/0411149*.

Rias, R. M., & Zaman, H. B. (2010, June). Investigating the redundancy effect in multimedia learning on a computer science domain. In *Information Technology (ITSim), 2010 International Symposium in* (Vol. 1, pp. 1-4). IEEE.

Richardson, J. T. (1987). Phonology and reading: the effects of articulatory suppression upon homophony and rhyme judgements. *Language and Cognitive Processes*, *2*(3-4), 229-244.

Richardson, J. T. (1996). Evolving concepts of working memory. *Working memory and human cognition*, 3-30.

Richardson, J. T., Engle, R. W., Hasher, L., Logie, R. H., Stoltzfus, E. R., & Zacks, R. T. (1996). *Working memory and human cognition*. Oxford University Press.

Rouinfar, A., Agra, E., Larson, A. M., Rebello, N. S., & Loschky, L. C. (2014). Linking attentional processes and conceptual problem solving: visual cues facilitate the automaticity of extracting relevant information from diagrams. *Frontiers in psychology*, *5*.

Rouinfar, A., Ph.D. thesis, Kansas State University, 2014.

Sadaghiani, H. R. (2011). Using multimedia learning modules in a hybrid-online course in electricity and magnetism. *Physical Review Special Topics-Physics Education Research*, *7*(1), 010102.

Salim, W., Setiawan, A. S., & Purnama, Y. (2012). Three-Dimensional Web-Based Physics Simulation Application For Physics Learning Tool. *Communication and Information Technology Journal*, *6*(2).

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010a). A closer look at split visual attention in system-and self-paced instruction in multimedia learning. *Learning and Instruction*, *20*(2), 100-110.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010b). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology*, *24*(2), 226-237.

Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, *19*(4), 469-508.

Smith, S. S. (2006). *Web-based instruction: A guide for libraries*. American Library Association.

Sohl, G (1989). *Het verwerken van de vreemdtalige gesproken tekst in een ondertiteld TV-programma* [Processing a foreign spoken text in a subtitled television program]. Unpublished thesis, University of Leuven, Leuven.

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport*,*11*(9), 2057-2061.

Stevenson, L. M., & Carlson, R. A. (2003). Information acquisition strategies and the cognitive structure of arithmetic. *Memory & Cognition*, *31*(8), 1249-1259.

Stiller, K. D., Freitag, A., Zinnbauer, P., & Freitag, C. (2009). How pacing of multimedia instructions can influence modality effects: A case of superiority of visual texts. *Australasian Journal of Educational Technology*, *25*(2), 184-203.

Stork, S., Stossel, C., & Schubo, A. (2008, August). Optimizing human-machine interaction in manual assembly. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 113-118). IEEE.

Swanson, H. L., Jerman, O., & Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *Journal of Educational Psychology*, *100*(2), 343.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive science, 12(2), 257-285

Sweller, J. (1999). Instructional Design in Technical Areas. Australian Education Review, No. 43. PCS Data Processing, Inc., 360 W. 31st, New York, NY 10001.

Tabbers, H. K., Martens, R. L., & Merriënboer, J. J. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, *74*(1), 71-81.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic bulletin & review*, *14*(4), 663-668.

Toh, S. C., Munassar, W. A. S., & Yahaya, W. A. J. W. (2010). Redundancy effect in multimedia learning: A closer look. *Curriculum, Technology & Transformation For An Unknown Future. Proceedings Ascilite Sydney*, 988-998.

Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology Section A*,*57*(1), 165-182.

Vallar, G., & Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), Handbook of memory disorders (2nd ed., pp. 249–270). Chichester: Wiley.

Van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior*,*25*(2), 325-331.

Van Gog, T., Paas, F., & Van Merriënboer, J. J. (2005). Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, *19*(2), 205-221.

Volkmann, F. C. (1986). Human visual suppression. *Vision Research, 26*(9), 1401-1416.

Weinstein, L. F., & Wickens, C. D. (1992). Use of nontraditional flight displays for the reduction of central visual overload in the cockpit. *The International Journal of Aviation Psychology*, *2*(2), 121-142.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical issues in ergonomics science*, *3*(2), 159-177.

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 360-380.

Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*, *105*(2), 266.

Appendix A - Problems Investigated

Table 5-1 The Skier task set

Initial Task	Rank the potential energy <u>lost</u> during the skier's descent down each slope from greatest to least. (That is, rank the potential energy <u>lost</u> from the start of A to the end of A vs. the start of B to the end of B vs. the start of C to the end of C; not the total value of potential energy.)
Training Tasks	Rank the potential energy <u>lost</u> during the skier's descent down each slope from greatest to least. (That is, rank the potential energy <u>lost</u> from the start of A to the end of A vs. the start of B to the end of B vs. the start of C to the end of C; not the total value of potential energy.)

Training Tasks	Rank the potential energy <u>lost</u> during the skier's descent down each slope from greatest to least. (That is, rank the potential energy <u>lost</u> from the start of A to the end of A vs. the start of B to the end of B vs. the start of C to the end of C; not the total value of potential energy.) Section A Section B Section C
	Rank the potential energy <u>lost</u> during the skier's descent down each slope from greatest to least. (That is, rank the potential energy <u>lost</u> from the start of A to the end of A vs. the start of B to the end of B vs. the start of C to the end of C; not the total value of potential energy.) Section A Section B Section C

| Training Tasks | Rank the potential energy <u>lost</u> during the skier's descent down each slope from greatest to least. (That is, rank the potential energy <u>lost</u> from the start of A to the end of A vs. the start of B to the end of B vs. the start of C to the end of C; not the total value of potential energy.)

Section A
Section B
Section C

Rank the potential energy <u>lost</u> during the skier's descent down each slope from greatest to least. (That is, rank the potential energy <u>lost</u> from the start of A to the end of A vs. the start of B to the end of B vs. the start of C to the end of C; not the total value of potential energy.)

Section A
Section B
Section C |

Training Tasks	Rank the potential energy <u>lost</u> during the skier's descent down each slope from greatest to least. (That is, rank the potential energy <u>lost</u> from the start of A to the end of A vs. the start of B to the end of B vs. the start of C to the end of C; not the total value of potential energy.) Section A Section B Section C
Near Transfer Task	A skier moves down a track. Rank the potential energy <u>lost</u> by the skier in each section of the track from greatest to least. The dotted lines indicate the beginning and ending of each section of the track. (Rank the potential energy <u>lost</u>, not the total value of the potential energy.) B A C

	A young girl slid down four frictionless playground slides as shown below. Compare the potential energy <u>lost</u> by the girl in each slide.
Far Transfer Task	

Table 5-2 The Ball task set

|	Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?
Initial Task	

	Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A? t = 0 s t = 1 s t = 2 s t = 3 s t = 4 s Ball A t = 0 s t = 1 s t = 2 s t = 3 s t = 4 s Ball B
Training Tasks	Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A? t = 0 s t = 1 s t = 2 s t = 3 s t = 4 s Ball A t = 0 s t = 1 s t = 2 s t = 3 s t = 4 s Ball B

Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?

t = 0 s t = 1 s   t = 2 s   t = 3 s   t = 4 s

Ball A

t = 0 s   t = 1 s   t = 2 s t = 3 s t = 4 s

Ball B

Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?

t = 0 s   t = 1 s   t = 2 s   t = 3 s   t = 4 s

Ball A

t = 0 s   t = 1 s   t = 2 s   t = 3 s   t = 4 s

Ball B

Training Tasks	Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?
	Two balls roll along the paths shown. A snapshot of the position of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?

Near Transfer Tasks	Ball A begins riding downward in an elevator at the same time Ball B is dropped from the roof of an adjacent building. A snapshot of the balls is taken every second. At what point in time does Ball B have the same speed as Ball A?
Far Transfer Task	A runner runs along a track. The following diagram, viewed from above, shows the position of the runner at each second. At which point in time is the runner moving the fastest?
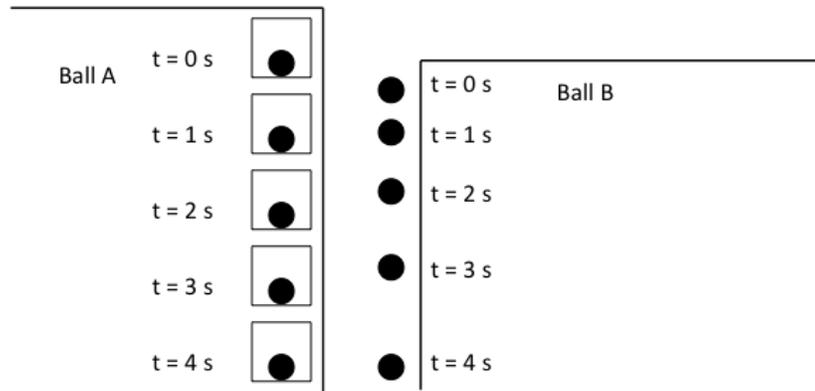
148
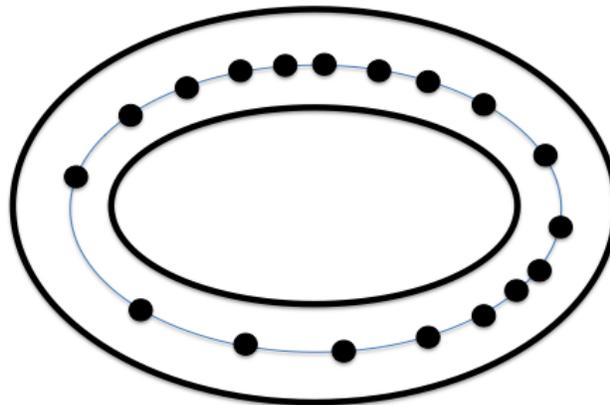
Table 5-3 The Roller Coaster task set

Initial Task	How does the final speed of cart A compare to the final speed of cart B, if the mass of the carts is the same and they both start at rest? (Frictional effects can be ignored)
Training Tasks	How does the final speed of cart A compare to the final speed of cart B, if the mass of the carts is the same and they both start at rest? (Frictional effects can be ignored)

|	
	How does the final speed of cart A compare to the final speed of cart B, if the mass of the carts is the same and they both start at rest? (Frictional effects can be ignored) A B Initial Final Initial Final _____ _____
Training Tasks	How does the final speed of cart A compare to the final speed of cart B, if the mass of the carts is the same and they both start at rest? (Frictional effects can be ignored) A B Initial Final Initial Final _____ _____

| Training Tasks	How does the final speed of cart A compare to the final speed of cart B, if the mass of the carts is the same and they both start at rest? (Frictional effects can be ignored)
	How does the final speed of cart A compare to the final speed of cart B, if the mass of the carts is the same and they both start at rest? (Frictional effects can be ignored)

Training Tasks	How does the final speed of cart A compare to the final speed of cart B, if the mass of the carts is the same and they both start at rest? (Frictional effects can be ignored) A Initial ... Final B Initial ... Final
Near Transfer Task	Two identical balls roll down a hill. How does the final speed of ball A compare to the final speed of ball B if the masses are the same and they both start at rest? (Frictional effects can be ignored) Initial A Final Initial B Final
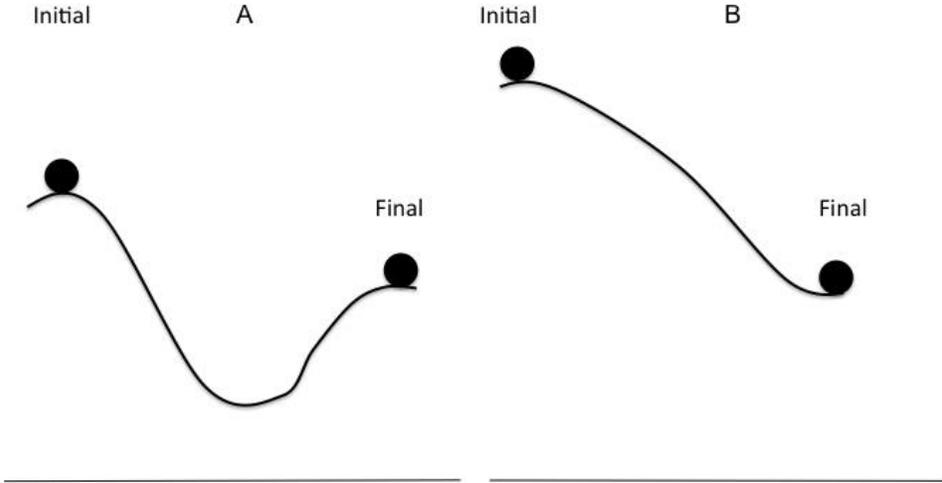
|	
Far Transfer Task	Two identical stones, A and B, are shot from a cliff with identical initial speeds. How do the speeds of the stones compare the instant before they hit the ground? A B
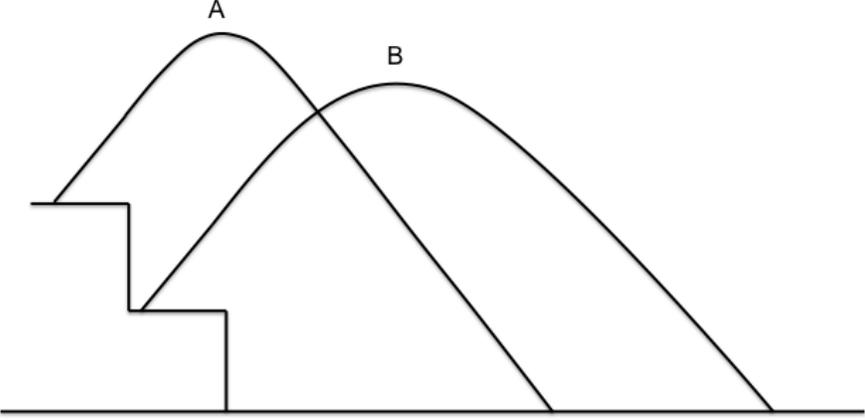
Table 5-4 The Graph task set

Initial Task	The motion of two objects is represented in the graph. When are the two objects moving with the same speed? position time
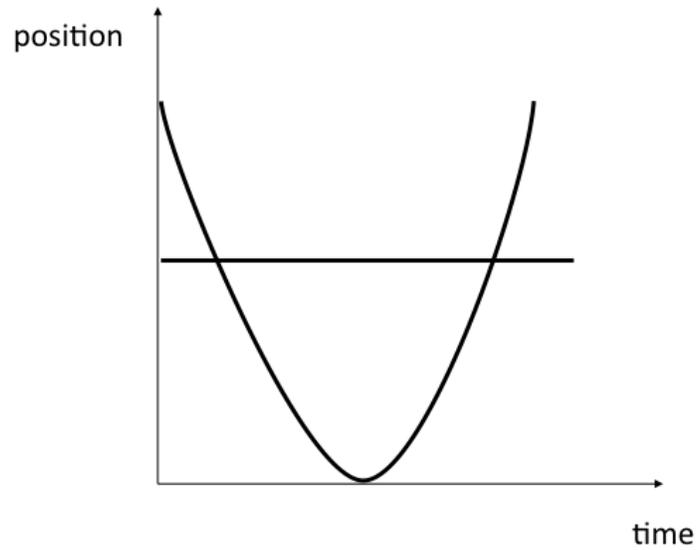
| Training Tasks | The motion of two objects is represented in the graph. When are the two objects moving with the same speed?

The motion of two objects is represented in the graph. When are the two objects moving with the same speed?

 |

|	
Training Tasks	The motion of two objects is represented in the graph. When are the two objects moving with the same speed? position time --- The motion of two objects is represented in the graph. When are the two objects moving with the same speed? position time

Training Tasks	The motion of two objects is represented in the graph. When are the two objects moving with the same speed?
	The motion of two objects is represented in the graph. When are the two objects moving with the same speed?

Near Transfer Task	The motion of two objects is represented in the graph. When are the two objects moving with the same speed? position time
Far Transfer Task	The motion of an object is represented by the graph. When is the object moving the fastest? position time
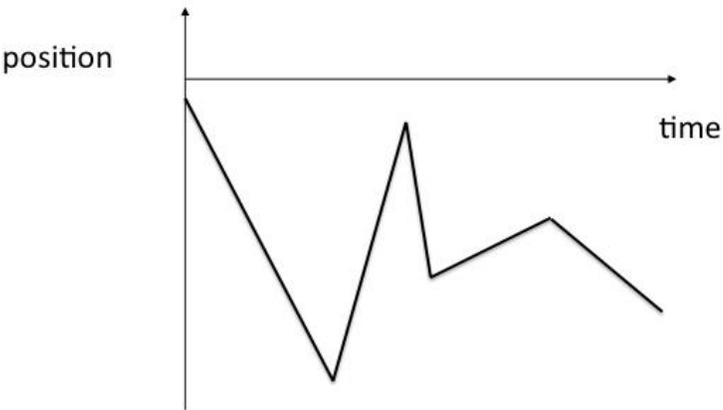
Appendix B - Results of the Mixed ANOVA on Dwell Time

In this supplemental material, we present a group of analyses on dwell time on visual hint AOI and text hint AOI. The type of analyses was identical to the corresponding analyses we have done on the domain relative ratios. In our opinion, normalizing the dwell time by using the total viewing time and the size of AOI is reasonable. But many eye movement studies used dwell time directly as the measure of visual attention.

A two-way 4 (conditions with visual hints, *i.e.*, visual condition, visual + audio condition, visual + text condition, and visual + text + audio condition) × 6 (six training tasks) mixed ANOVA with repeated measures on training tasks was conducted to compare dwell time on visual hint AOI during hints. There was a significant main effect of training tasks, $F(5, 315) = 5.700, p < .001$. And there was a significant main effect of conditions, $F(3, 63) = 11.819, p < .001$. These two main effects were qualified by the significant interaction, $F(15, 315) = 3.219, p < .001$ (see Figure 5-1). To explore participants' visual attention change in the whole training process, dwell times were compared across all six training tasks for each condition. Four one-way repeated measures ANOVA were conducted. We found a significant effect of training tasks in the visual condition, $F(5, 70) = 9.080, p < .001$. There is no significant effect of training tasks in visual + audio condition, $F(5, 80) = 2.410, p = 0.061$; visual + text condition, $F(5, 70) = 1.107$, $p = 0.361$; and visual + text + audio condition, $F(5, 95) = 0.080, p = 0.990$.

To further probe the interaction, in order to compare the way that hint modalities affect participants' visual attention on each training task, dwell times were compared across four conditions in each training task. Six one-way ANOVA were conducted. All of them showed significant effects of conditions (see Table 5-5). Post hoc pairwise comparisons with Bonferroni correcting found the significant differences occurred only between the conditions with text hints

158

(i.e., visual + text and visual + text + audio) and the conditions without text hints (i.e., visual and

visual + audio). There is no significant difference for adding audio hints or not (see Table 5-6).

Another two-way 4 (conditions with text hints, *i.e.,* text condition, text + audio condition,

visual + text condition, and visual + text + audio condition) × 6 (six training tasks) mixed

ANOVA with repeated measures on training tasks was conducted to compare domain relative

ratio on text hint AOI during hints. There was a significant main effect of training task, $F(5, 340)$

$= 47.232$, $p < .001$, and a significant main effect of condition, $F(3, 68) = 23.805$, $p < .001$. There

was also a significant interaction, $F(15, 315) = 1.784$, $p = 0.049$ (see Figure 5-2). To explore

participants' visual attention change in the whole training process, dwell times were compared

across all six training tasks for each condition. Four one-way repeated measures ANOVA were

conducted. We found a significant effect of training tasks in the text condition, $F(5, 95) =$

13.353, $p < .001$; visual + text condition, $F(5, 70) = 9.403$, $p < .001$; text + audio condition, $F(5,$

$80) = 19.764$, $p < .001$; and visual + text + audio condition, $F(5, 95) = 9.902$, $p < .001$.

To further probe the interaction, in order to compare the way that hint modalities affect

participants' visual attention on each training task, dwell time were compared across four

conditions in each training task. Six one-way ANOVA were conducted. All of them showed

significant effects of conditions (see Table 5-7). Post hoc pairwise comparisons with Bonferroni

correcting found the significant differences occurred only between the conditions with visual

hints (i.e., visual + text and visual + text + audio) and the conditions without visual hints (i.e.,

text and text + audio). There is no significant difference for adding audio hints or not (see Table

5-8).

To probe how participants split their visual attention when visual hints and text hints

were presented simultaneously, a two-way 2 (visual + text condition and visual + text + audio

condition) × 2 (visual hint AOI and text hint AOI) mixed ANOVA with repeated measures on

AOI was conducted (see Figure 5-3). There was no main effect of condition, $F(1, 33) = 0.107$, $p = 0.746$, which means adding audio hints or not does not affect participants' split of their visual

attention between visual hints and text hints. There was a significant main effect of AOI, $F(1, 33) = 9.812$, $p = .004$. The dwell time on visual hints is significantly longer than on text hints

(1712.9 ± 87.7 vs. 1251.1 ± 101.3, respectively)

Figure 5-1 The dwell time on the visual hint AOI on the training tasks. Error bars indicate ±1 standard error of the mean.
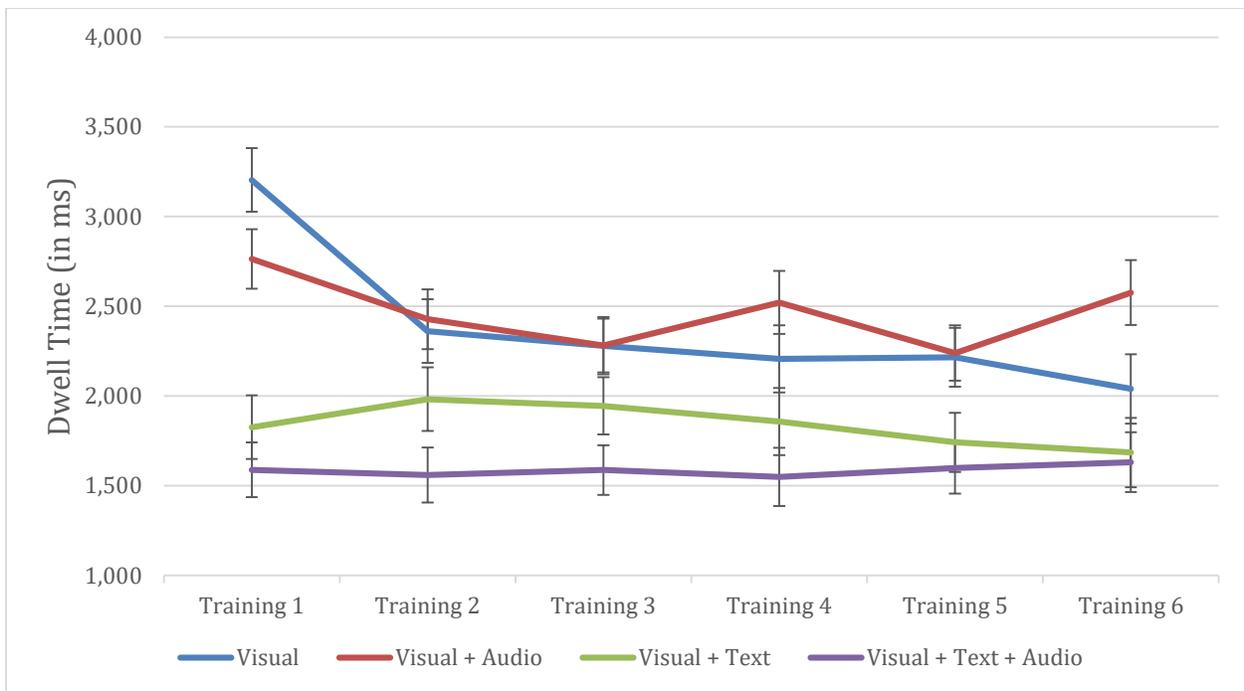

Table 5-5 The *F* values and *p* values of comparing the dwell time on the visual hint AOI between visual condition, visual + text condition, visual + audio condition, and visual + text + audio condition for each training task.

Training Task	#1	#2	#3	#4	#5	#6
F	20.872	6.171	5.219	6.114	4.600	5.879
p	< .001	0.001	0.003	< .001	0.006	0.001

Table 5-6 The *p* values from pairwise comparisons with Bonferroni correction on the dwell time on visual hint AOI between conditions for each training task.

Training Task	#1	#2	#3	#4	#5	#6
Visual vs. V + T	< .001	0.817	0.853	1.000	0.272	1.000
Visual vs. V + A	0.447	1.000	1.000	1.000	1.000	0.280
Visual vs. V + T + A	< .001	0.007	0.010	0.059	0.036	0.686
V + T vs. V + A	0.002	0.435	0.780	0.073	0.184	0.008
V + T vs. V + T + A	1.000	0.459	0.568	1.000	1.000	1.000
V + A vs. V + T + A	< .001	0.002	0.007	0.001	0.020	0.002

Figure 5-2 The dwell time on the text hint AOI on the training tasks. Error bars indicate ±1 standard error of the mean.
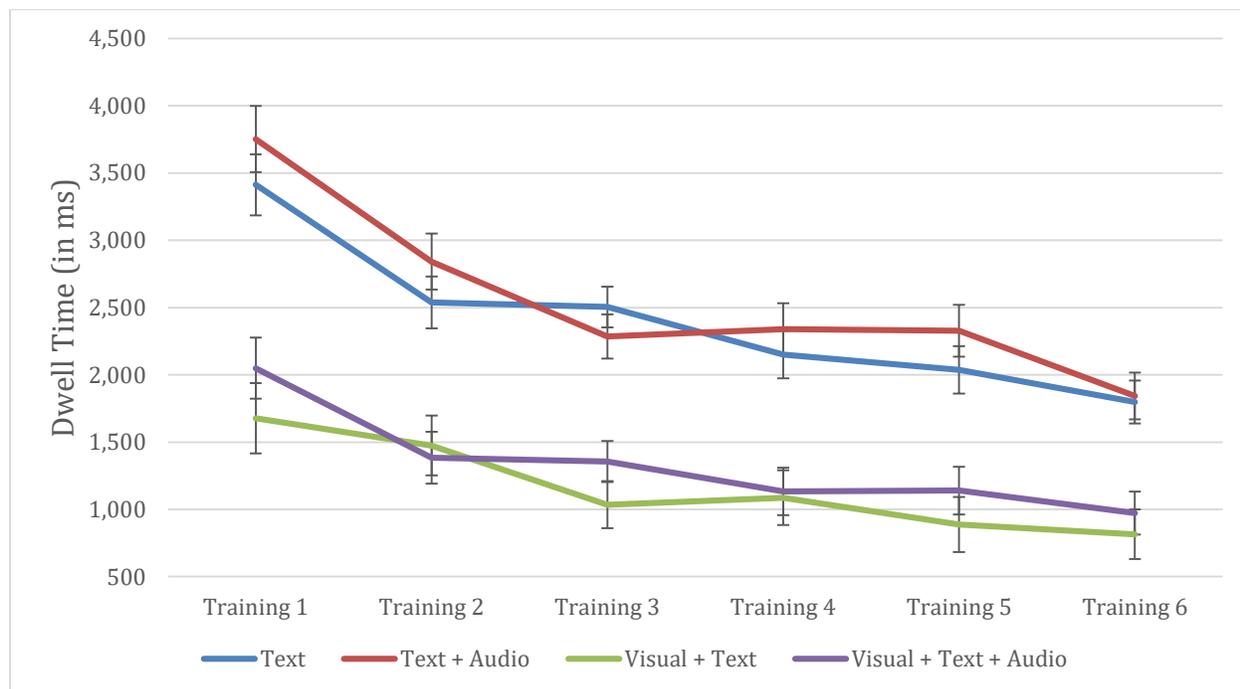

Table 5-7 The *F* values and *p* values of comparing the dwell time on the text hint AOI between visual condition, visual + text condition, visual + audio condition, and visual + text + audio condition for each training task.

Training Task	#1	#2	#3	#4	#5	#6
F	17.108	13.127	19.252	12.404	13.119	9.918
p	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$

Table 5-8 The p values from pairwise comparisons with Bonferroni correction on the dwell time on text hint AOI between conditions for each training task.

Training Task	#1	#2	#3	#4	#5	#6
Text vs. V + T	$< .001$	0.003	$< .001$	0.001	$< .001$	0.001
Text vs. T + A	1.000	1.000	1.000	1.000	1.000	1.000
Text vs. V + T + A	$< .001$	$< .001$	$< .001$	0.001	0.004	0.003
V + T vs. T + A	$< .001$	$< .001$	$< .001$	$< .001$	$< .001$	0.001
V + T vs. V + T + A	1.000	1.000	1.000	1.000	1.000	1.000
T + A vs. V + T + A	$< .001$	$< .001$	0.001	$< .001$	$< .001$	0.003

Figure 5-3 The dwell time on the text hint AOI and the visual hint AOI. Error bars indicate ±1 standard error of the mean.