Benefits of multisensory presentation on perception, memory and navigation

ISBN/EAN: 978-90-393-5751-4

Benefits of multisensory presentation on perception, memory and navigation

De voordelen van multisensorische presentatie
op perceptie, geheugen en navigatie
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag
van de rector magnificus, prof.dr. G.J. van der Zwaan, ingevolge het besluit
van het college van promoties in het openbaar te verdedigen op

woensdag 21 maart 2012 des ochtends te 10.30 uur

door

Tom Gijsbert Philippi

geboren op 19 juni 1983, te Rhenen

Promotor: Prof.dr. P.J. Werkhoven
Co-promotor: Dr. J.B.F. van Erp

Table of contents

Chapter 1. Introduction

Abstract

Navigation is the process of planning and following routes to travel from the current location to a target location. In comparison with real world navigation, we have considerable difficulty with navigation in virtual environments. An important cause is that less information is presented in a virtual environment than in the real world. For example, virtual environments are often limited to visual presentation only. Currently, little is known how adding auditory and tactile information to virtual environments affects navigation performance. The presentation of information about a single object or event to multiple sensory modalities can be advantageous to perception and memory. Because navigation largely relies on both perception (e.g. where am I?) and memory (e.g. where do I need to go and how do I get there?), multisensory benefits may also apply to navigation. Adding auditory and tactile information may therefore improve navigation in (visual) virtual environments.

In this chapter, we introduce the Trisensory Orientation Model (TOM). TOM models navigation, its dependence on perception and memory, and the potential benefits of congruent presentation of visual, auditory and tactile information on each of these processes. Congruent refers to the presentation of information in multiple sensory modalities at the same time, from the same place and with the same meaning. With our model in mind, we formulate seven research questions and hypotheses which are investigated in this thesis. These questions are concerned with the underlying mechanisms of visuotactile interactions, the role of stimulus congruency in perceptual task performance, the effects of visuotactile presentation on the encoding and retrieval of object identity and object location memory, the multisensory memory effects for congruent semantic and non-semantic items, and the benefits of audiovisual landmark presentation on spatial memory and navigation performance.

1.1 Navigation

One of the most fundamental activities in our daily life is our interaction with the environment. An important element of this interaction involves movement through space. A difference can be made between simple locomotion, i.e., the physical displacement through space, and navigation, which involves complex cognitive processes such as building spatial representations.
Navigating is the cognitive process of planning and following routes to travel from the current location to a target location. We rely on navigation to find our way in complex environments. When navigating, a navigator is concerned with the answers to the questions such as: Where am I?, where do I need to go? and how do I get there? These questions are mostly resolved using perception and memory. For instance, we can compare the surroundings we perceive to our memories of previously visited locations to determine where we are, to determine if we are still following the route we planned, and to determine if we have arrived at our intended destination.

Navigation in the real world
When we navigate an environment, we learn about that environment. We receive (spatial) information from multiple sensory cues which we can store and use to guide subsequent navigation behavior. Earlier literature suggests a distinction between three different types of information relevant to the navigator (Siegel & White, 1975; Thorndyke, 1981). These are landmark, route and configurational information. Landmark information refers to knowledge about the appearance of objects that stand out in the environment. Landmarks serve as reference points and are helpful to determine where we are (Janzen, 2006). Route information refers to the knowledge about the sequence of actions required to follow a route between two locations. It contains information about when a navigator has to change direction when following a route, similar to the information you would receive when you ask for directions. As such, route information may make use of landmark information. Configurational information refers to topological (e.g. map-like) knowledge about the locations of landmarks and the spatial relations between these locations. This type of information allows us to take shortcuts (Tolman, 1948)
In real world navigation, the components of a landmark may be visual, but could also be auditory or tactile. These may all aid us during navigation. Furthermore, it has been shown that perceiving information about an object or event with multiple modalities often benefits performance. For instance, multisensory presentation can shorten reaction time (Bernstein, Clark & Edelstein, 1969; Hershenson, 1962; Nickerson, 1973), enhance stimulus detection (Vroomen & De Gelder, 2000), and enhance our memory (Thompson & Paivio, 1994; Shams & Seitz, 2008). We think that these benefits may also apply during navigation.

Navigation in virtual environments
Navigation is not limited to real worlds. Nowadays, computer systems allow us to create virtual environments with relative ease. Virtual environments may range from visualizations of simple file directories to three-dimensional simulations of complex

physical places (or even imaginary ones). The latter are often used for our entertainment, but have also been used as an alternative to real world training, and for product design and prototyping (Bowman, 1995).

We are interested in simulations of complex places because they are a very promising as an alternative to real world training. The potential advantages of using them for training are cost (e.g., a flight simulator is much cheaper than flying a real jet), customizability (e.g., weather can be changed on command), and safety (i.e., flying a jet is much more dangerous than flying in flight simulator). In this thesis, we will use the term virtual environments only to refer to three-dimensional simulations of complex places.

Differences and similarities of navigation in real and virtual environments

The experience in a virtual environment differs from that in the real world. Interactions with virtual environments are mostly visual experiences, displayed on a computer screen. This computer screen has a limited resolution and a limited field of view and the environment that is displayed typically has less detail than a real world environment. Furthermore, movement control in virtual environments may differ from that in a real world. For example, in most virtual environments it is not possible to control movement by physically walking (but see Souman *et al.,* 2011) Each of these differences may result in less information being available to the navigator in virtual than in real environments.

In spite of these differences, we learn landmark, route and configurational information when exploring virtual environments (Giller & Mallot, 1998; Richardson, Montello & Hegarty, 1999). However, we can experience considerably difficulty with navigation in virtual environments (Ellis, 1993; Kaur, Sutcliffe & Maiden, 1999). For instance, people have reported getting lost more easily in virtual than in real environments (Wilson, Foreman & Tlauka, 1997). Furthermore, we learn (spatial) information about a virtual environment more slowly than about a real one (Richardson *et al.*, 1999).

The difficulty experienced with the navigation of virtual environments is the result of the availability of less information than in a real world. Increasing the amount of visual information can improve navigation performance in virtual environments (Riecke, Van Veen & Bülthoff, 2002). Others, however, have argued that much more can be gained by presenting information in other sensory modalities (Ruddle & Lessels, 2006). Bakker (2001), for instance, shows that the presentation of visual, kinaesthetic and proprioceptive information can improve navigation in virtual environments in comparison with the presentation of only visual information.

Presenting multisensory information

Several differences can be made in how information is presented in multiple sensory modalities. The information presented in each modality may either be about the same or about a different object or event. While large benefits on navigation in virtual environments have been reported when auditory (Burrigat & Chittaro, 2007; Gunther, Kazman & MacGregor, 2004) or tactile (van Erp, 2007) information was about different objects and events than the presented visual information, little is known, however, about any (beneficial) effects of presenting visual, auditory and tactile information about the same objects and events.

When the information presented to multiple modalities is about the same object or event it can either be redundant or complementary. When multiple sensory modalities measure different dimensions they complement each other. In contrast, when multiple sensory modalities measure an object or event in the same dimension, those measurements are redundant. Redundant measurements can either be congruent (i.e. each sense returns the same value) or incongruent (i.e. each sense returns a different value). When we perceive redundant information about the objects and events in the real world with multiple sensory modalities, that information is often congruent.

Purpose of this thesis

The overarching purpose of the present thesis is to explore the possible benefits of congruent visual, auditory and tactile information in navigation. Because navigation largely relies on perception and memory we expect that the multisensory benefits reported for those two processes also apply to navigation. These benefits may be relevant for the (design of) training and other applications employing virtual environments.

1.2 Models of navigation

In this section we further explore how multisensory presentation may affect navigation. We start with describing two models, viz. Wickens' general model of information processing (Wickens, 1992; Wickens & Holland, 2000) and the Framework for Investigation of Navigation and Disorientation (FIND; Bakker, 2001). Both models have a loop character where the perception of an environment (or stimulus in that environment) leads to an action which in turn may lead to a new perception (see also Boyd's Observe, Orient, Decide, and Act-loop; Boyd, 1987; Osinga, 2006).

Both models are capable of processing multisensory information but do not make predictions about the effects of congruent visual, auditory and tactile presentation on navigation in virtual environments. Therefore, we introduce the Trisensory Orientation Model (TOM). This is a variant of FIND that models the visual, auditory and tactile interactions relevant to navigation. The potential benefits of multisensory presentation are discussed with this model in mind.

Wickens' model of human information processing and memory

Wickens proposed a general model for human information processing and memory, which is illustrated in Figure 1.1. The model describes how sensory stimuli are processed and lead to a response. Firstly, a stimulus is received which (affected by memory and attention) results in a percept. After perception, an appropriate response is made on the basis of allocated attention and previous experience (which is retrieved from memory). Finally, the response is executed, which affects the environment and may, in turn, affect the next stimulus that is received.

Wickens' model shows that sensory information is processed in multiple sensory modalities. However, neither the effects of multisensory processing nor the process of navigation are made explicit. Thus, the model is also not explicit about multisensory effects on navigation.

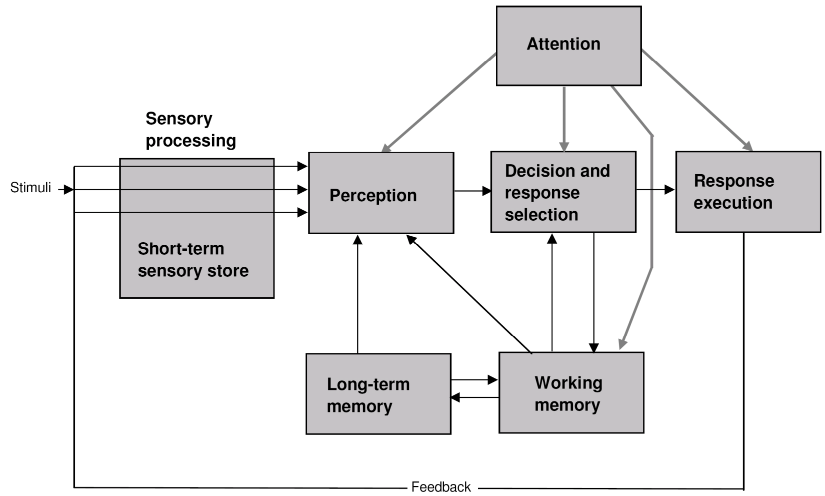


Figure 1.1: Wickens' model for human information processing and memory (adapted from Wickens, 1992). Stimuli from the environment are received and processed, which results in perception. On the basis of perception, attention, and memory a response is selected and executed which in turn may affect the next stimulus that is received.

Framework for the investigation of navigation and disortientation

Figure 1.2 depicts FIND a variant of Wickens' model of information processing (Bakker, 2001, p. 4-9) that specifically models the process of navigation in virtual environments and the potential effects of visual, vestibular and kinaesthetic stimulation on navigation. In FIND a virtual environment generates stimuli that are perceived by the human body. These stimuli are then interpreted by *path integration* and *recognition,* and stored in the *cognitive map*. *Path integration* deals with translating sensory cues into an estimation of displacement, while *recognition* is concerned with identifying objects that have been encountered earlier. These objects, as well as the configuration of the virtual environment, are stored in the *cognitive map*. *Cognitive anticipation* is the expectancy to arrive at a certain location when following a route. It was included in FIND to study the effect of discontinuous displacements (i.e. teleporting) in virtual environments. Based on *recognition*, *path integration*, and knowledge about the environment the current location of the navigator is determined. Movement is then determined and executed, based upon the current location and the navigator's knowledge of the environment (which is stored in the *cognitive map*). The processes in FIND are controlled by a *cognitive control system*. This system determines to which processes resources (such as attention or short term memory) are allocated in order to achieve the goal.

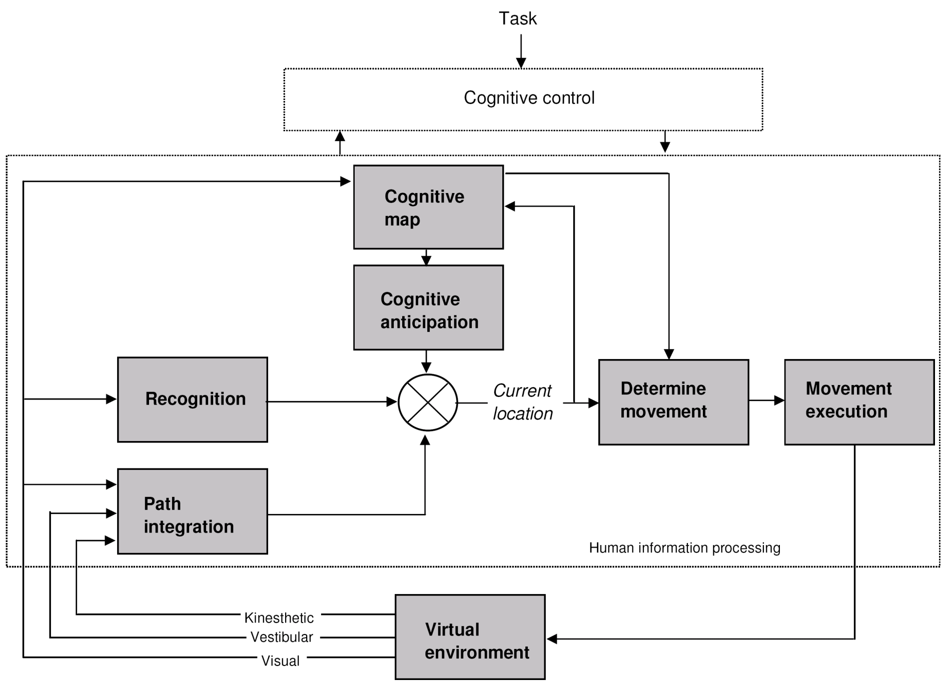


Figure 1.2: Framework for the Investigation of Navigation and Disorientation (adapted from Bakker, 2001). The virtual environment stimulates the kinaesthetic, vestibular and visual senses of the body. From these sensory signals a current location is determined by employing path integration, recognition and a cognitive map about the environment. Based upon the current location, and knowledge about the environment from the cognitive map movement is determined and executed, which affects the body and the virtual environment.

FIND explicitly describes the process of navigation and is therefore highly relevant to this thesis. In addition, FIND models multisensory effects for visual, vestibular and kinaesthetic information presentation on *path integration*. Improved path integration subsequently improves the *cognitive map* and aids movement decisions. However, FIND does not model auditory and tactile information processing and therefore also does not make explicit the possible effects of congruent visual, auditory and tactile presentation.

A navigation model processing congruent visual, auditory and tactile information

Both the models we just described are capable of processing information from multiple sensory modalities. However, both models do not specify if and how congruent presentation of vision, audition and touch may affect subsequent navigation behavior. Yet, multisensory research indicates that such presentation should at least affect perception (Ernst & Bülthoff, 2004) and memory (Shams & Seitz, 2008).

We present a navigation model, which incorporates these findings. Figure 1.3 depicts TOM (Trisensory Orientation Model), an adaptation of FIND that focuses on the processing of visual, auditory and tactile information. The single purpose of TOM is to illustrate how congruent visual, auditory and tactile information may affect navigation in virtual environments. Predictions about and interactions with other than these sensory modalities are therefore not included. TOM is a qualitative description of the processes involved and does not result in any quantitative predictions. The modules in TOM and the (possible) effects of multisensory presentation on them are discussed in the next pages.

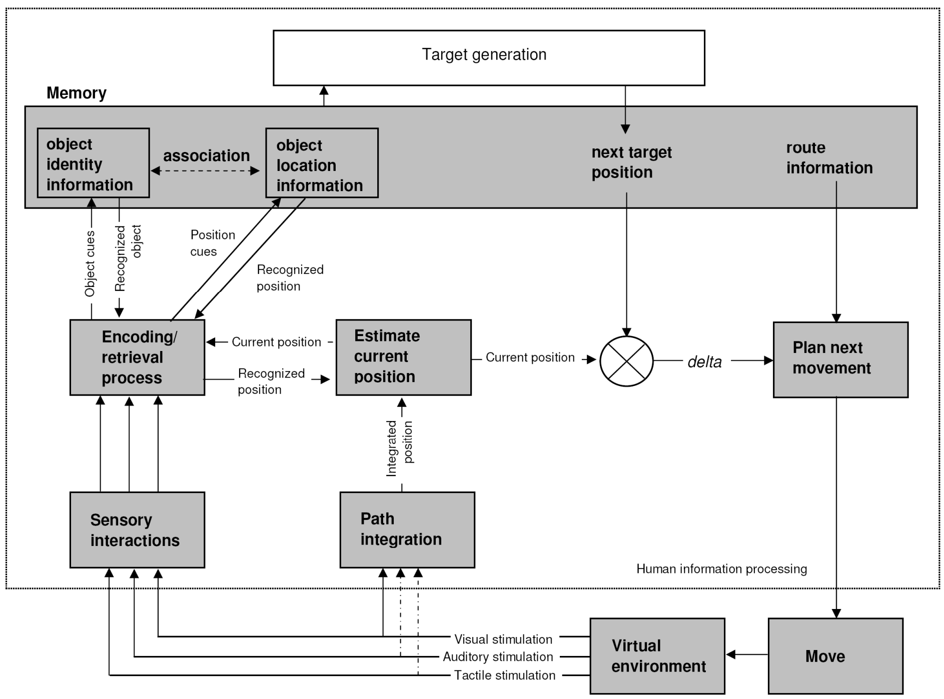


Figure 1.3: Trisensory Orientation Model (TOM). TOM assumes a sequence of processes at the perceptual, memory and navigation level and describes how they may be affected by congruent visual, auditory and tactile presentation. First, a navigator determines where he wants to go (i.e. *target generation* sets a *next target position*). The navigator then attempts to match his current position with the target position. This works as follows. The navigator perceives sensory information from the environment. Some of this information (i.e., mainly visual flow information) is processed in *path integration* to determine displacement, which helps the navigator to determine his *current position*. Information perceived with multiple modalities that is about the same object or event is affected by *sensory interactions*. *Encoding/retrieval processes* then encode this information and compare it with *memory*. A successful match with *memory* allows objects and their locations to be recognized, which may aid the navigator to estimate his *current position*. The estimate of the *current position* is compared with a *target position*; a difference between these positions means that the navigator has not yet reached his target. Based upon this difference, and upon route information in *memory*, movement is planned and executed which affects the navigator's position and orientation in the virtual environment.

Path integration

In *path integration* sensory information is used to determine displacement. It has been shown that path integration mainly depends on visual flow and proprioceptive information (Warren, Morris & Kalish, 1998; Bakker, 2001). However, it is conceivable that sound and touch could improve path integration. For instance, when driving a car, the sound of the engine could give the driver a hint about the car's speed, which may contribute to the driver's estimate of the car's displacement. Therefore, in TOM, visual, auditory and tactile information contribute to *path integration*. Note though, that path integration and the possible contributions of auditory and tactile information to it are not investigated in the present thesis.

Sensory interactions

Humans, like most other organisms, have a variety of sensory modalities through which they perceive their surroundings (Stein & Meredith, 1993). Each of the sensory modalities provides a distinct impression of our environment. Traditionally, the senses were considered to largely operate as separate and independent processes (Boring, 1942) and, as a consequence, they were mostly studied in isolation.

However, sensory information from the senses is not processed in isolation, but automatically combined into a single, holistic percept (Ernst & Bülthoff, 2004). For instance, when someone speaks to us, we do not perceive his or her lip movements and speech separately. Instead, lip movements and speech are integrated into a coherent whole and, perhaps more importantly, this allows lip movements to affect our perception of speech (see McGurk & MacDonald, 1976).

Research on the effects of multisensory presentation has revealed many sensory interactions. They have been reported at the neural level (Stein & Meredith, 1993) and the behavioral level in perception (Ernst & Bülthoff, 2004), attention (Spence & Driver, 2004), and memory (Shams & Seitz, 2008). These interactions often have favorable effects. In comparison with unisensory presentation, multisensory presentation can shorten reaction time (Bernstein, Clark & Edelstein, 1969; Hershenson, 1962; Nickerson, 1973), enhance stimulus detection (Vroomen & De Gelder, 2000), and increase the number of items encoded in memory (Thompson & Paivio, 1994).

In TOM, a virtual environment produces visual, auditory, and tactile sensory cues. The findings in multisensory research make clear that from the moment these cues are perceived, they may start to interact. In TOM, the perception of sensory cues and their interactions are captured by *sensory interactions*, which affect the quality of the processed signals and therewith the performance of subsequent processes. How subsequent processes are affected by *sensory interactions* is explained in the paragraphs about these processes.

When do sensory cues interact?

It is important to note that not all sensory cues that are perceived should interact with each other. That is, sensory cues should only do so when it is reasonable to assume that they originate from the same object or event (unity assumption; see Welch & Warren, 1980). Because it is *a priori* impossible to determine whether two sensory signals originate from the same source, several criteria are (automatically) employed to estimate whether sensory cues from different sensory modalities should interact. Figure 1.4 illustrates how this process may work.

Two important criteria are the spatial and temporal correspondence of sensory cues across modalities. Because two objects cannot occupy the same space at the same time it is likely that sensory cues originating from the same space at the same time are from a single object. Indeed, the effects of multisensory presentation at the neural and behavioral level are larger when that is roughly the case (Stein & Meredith, 1993; Shams, Kamitani, & Shimojo, 2002).

A third criterion which affects sensory interaction is semantic congruency. Different multisensory effects have been reported for sensory stimuli with the same and different semantic information (Laurenti *et al.* 2004; Lehmann & Murray, 2005). For instance, the response time to confirm that a picture depicts a cow decreases when a picture of a cow is simultaneously presented with a sound of a cow but not when that picture is presented with a (non-semantic) beep (Yuvall-Greenberg & Deoull, 2007). Obviously, determining semantic congruence requires prior experience. Sensory cues need to be identified before they can be compared. This means that the interactions following a semantic-congruency check may occur later than the interactions following spatial and temporal correspondence. In TOM, we assume that the spatial- and temporal-congruency checks take place prior to the *encoding/retrieval process* while the semantic-congruency check takes place during that process.

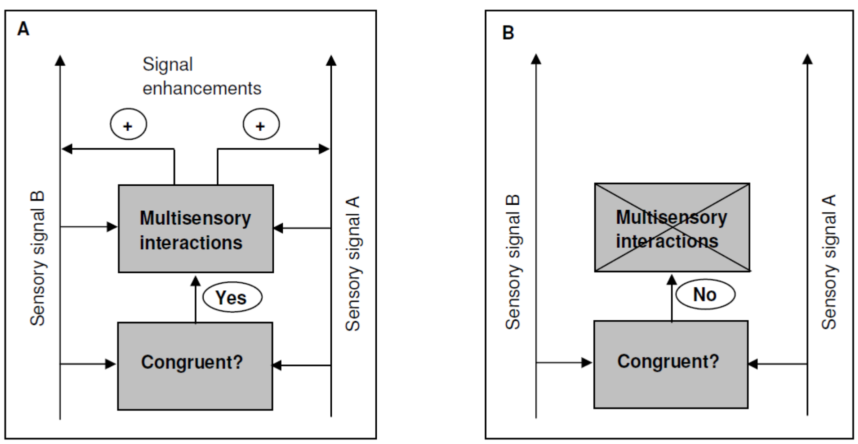


Figure 1.4. When do sensory cues interact? The two panels further illustrate the workings of *sensory interactions*. When two sensory cues from two different modalities are processed it is checked whether they are spatially, temporally, and/or semantically congruent. When they are (left panel), it is likely that they are about the same object and event and they should interact, which enhances the signals. When they are not (right panel), no interactions take place and the signals are processed separately[1].

[1]Note that Figure 1.4 is a simplification of the involved processes. Because perception is naturally noisy, multisensory stimuli that are physically congruent may be perceived as (slightly) incongruent. Consequently, stimuli that are physically incongruent are sometimes (mis)treated as congruent stimuli and subjected to sensory interactions (see Bresciani, Dammeier & Ernst, 2006; Shams, Kamitani & Shimojo, 2000, 2002).

Sensory combination and integration

Related to the above criteria are two principles of sensory interaction. Ernst and Bülthoff (2004) identify them as sensory combination and sensory integration. Sensory combination occurs when we perceive complementary information in multiple senses. In perception this has the potential to disambiguate a situation. A striking example of sensory combination is the motion-bounce illusion (Sekuler, Sekuler and Lau, 1997). In this illusion two circles are shown to move from the upper corners to the lower corners of a screen, passing through each other in the middle. Observers can perceive these circles as either crossing or bouncing. However, when a sound is presented when the circles pass through each other, many observers perceive the crossing circles as bouncing instead. Sekular and colleagues suggested that this qualitative change occurs because the sound is associated with a bouncing movement. In other words, the semantic information carried by the sound interacts with the semantic information carried by the crossing circles.

Sensory integration occurs when we perceive redundant information in multiple sensory modalities. The effect of sensory integration has often been investigated in studies with artificially created intersensory discrepancies. If these discrepancies are small, sensory processes still consider the presented information as congruent and integrate it. Shams and colleagues (2000; 2002), for instance, presented one flash together with two rapid beeps and found that the presentation of beeps affected the number of perceived flashes. Follow-up studies with presentation of larger numbers of flashes and beeps also argued that sensory integration reduced the variance in the number of perceived flashes. In other words, the presentation of redundant information can affect the mean and reliability (which is the inverse of the variance) of sensory estimates.

It is important to note that most objects and events we perceive in our environment carry both complementary and redundant information in different sensory modalities. This means that they may be subject both to sensory combination and to sensory integration.

Memory

Memory is the neurocognitive capacity to encode, store and retrieve information (Schacter, 2007). In TOM, *memory* holds all information relevant to the navigator. This includes information about the identity and location of landmarks, route and configuration information, and information about the navigator's current and target positions. We view *memory* as a database that only holds information. Access to and editing of the database is modeled by a separate *encoding and retrieval process*.

Studies on the processing of object identity (e.g. what?) and object location information (e.g. where?) indicate that object identities and object locations are processed separately in memory (Aquirre, D'Esposito, 1997; Lehnert & Zimmer, 2006; 2008; Moscovitch *et al.,* 1995). Therefore, we explicitly distinguish stored information related to object identities and object locations in *memory*. Furthermore, we assume that spatial processes transfer (or translate) information about locations between route and configuration information where appropriate (see Siegel & White, 1975; Ishiwaka & Montello, 2006). Finally, we assume an association between object identity and object location which helps us to remember an object identity when we think of an object location and vice versa (see Postma, Kessels & Van Asselen, 2004; 2008 for description of a similar mechanism).

Multisensory memory effects

In comparison with unisensory presentation, multisensory presentation improves our perception. Thus, it is obvious that multisensory presentation can affect memory performance. After all, we cannot remember an object that we did not (properly) perceive. A more interesting question is, therefore, if multisensory presentation improves memory even when the unisensory components are perfectly perceivable. Therefore, we assume (and ensure) perfect perception of unisensory properties when we investigate memory effects in this thesis.

Several studies reported beneficial effects of multisensory presentation on memory performance (Lehmann & Murray, 2006; Murray *et al.* 2005; Thompson & Paivio, 1994; Von Kriegstein & Giraud, 2006). Thompson and Paivio, for instance, showed that matching sounds and pictures were memorized 50% more often than sounds or pictures presented alone. Importantly, the sounds and pictures presented in that study could be easily identified, which means that the multisensory increase in memory performance was not due to multisensory presentation increasing perceptual performance.

Currently, there are no known effects of multisensory presentation on memory for locations. However, multisensory perception research has shown that sensory interactions also occur for object locations. Therefore, we expect that multisensory memory effects may also exist for object locations. In addition, because objects are associated with locations, it is also possible that if multisensory presentation makes an object identity easier to memorize that the associated location also becomes easier to memorize. In TOM, we therefore assume that multisensory interactions also affect memory for locations.

Encoding and retrieval

Adequate memory performance relies both on encoding information in memory and on retrieving information from memory. So far, the studies that reported multisensory memory benefits (Lehmann & Murray, 2006; Murray *et al.* 2005; Thompson & Paivio, 1994; Von Kriegstein & Giraud, 2006) have either assessed multisensory effects on encoding or have not distinguished between both processes. Studies investigating the relation between encoding and retrieval indicate that retrieval is the reactivation of encoding operations (Slotnick, 2004; Kent & Lambers, 2008). This means that these processes are largely similar. In TOM we therefore assume a single *encoding and retrieval process.* We think that multisensory effects reported for encoding also apply to retrieval.

Multisensory memory mechanisms

Generally speaking, there are two possible mechanisms how multisensory presentation may improve memory (see also Figure 1.5). Firstly, multisensory interactions may improve the effectiveness of encoding and retrieval. Multiple sensory modalities gather more (accurate) information than a single one and that may reduce the signal to noise ratio of the *encoding/retrieval process*. This effect should not be confused with sensory interactions improving perception. For instance, it is conceivable that a picture of an aircraft can be better memorized than another picture of the same aircraft even though both pictures clearly show an aircraft. The first picture may contain more detail and thus result in more effective encoding/retrieval. Multisensory presentations may likewise result in more effective encoding/retrieval than unisensory presentations because they provide more ‘detail’. Also note that this mechanism is subject to the ‘congruency checks’ we discussed earlier. That is, with this mechanism multisensory benefits occur only when the information presented in each sensory modality has the same timing, location and/or meaning.

Secondly, multisensory presentation may lead to multiple *encoding/retrieval processes*, one for each sensory modality. Memory performance is then increased because each additional sensory modality functions as a back-up. If one sensory modality fails to encode or retrieve certain information, the second may still be successful, increasing the effectiveness of multisensory memory over unisensory memory (probability summation). Note that with this second mechanism, multisensory effects are not subject to “congruency checks”. Memory performance may increase even when information is presented from different locations at different times and with different meanings.

A version of the second mechanism is described by the dual coding theory (Paivio, 1971; 1986). The dual coding theory assumes that modality-specific verbal and non-verbal components are processed and memorized independently. The presentation of a sound and picture thus results in visual and auditory memory representations, which can be accessed independently. The dual coding theory, however, is not explicit whether modality-specific retrieval processes can access information stored via another modality. This means that the dual coding theory does not predict an effect of multisensory presentation on memory retrieval after unisensory encoding. Studies on cross-sensory memory performance, however, show that modality-specific cues can be used to retrieve information that was presented earlier in another modality (Ernst, Lange & Newell, 2007; Woods & Newell, 2004; Butler & James, 2011). This means is the dual coding theory is not up to date. In Figure 1.5, we therefore illustrated a version of the second mechanism that allows modality-specific processes to access all information in memory.

It is not yet clear which of these two mechanisms is responsible for the benefits reported in multisensory memory. In fact, it is even possible that both processes provide multisensory memory benefits (in that case multiple enhanced encoding/retrieval processes would be initiated). In this thesis, we are first and foremostly interested in if there are multisensory benefits on memory and navigation. Therefore, we incorporated both mechanisms in TOM. On the one hand, sensory signals initiating encoding and retrieval are subject to *sensory interactions*. On the other hand, multiple sensory signals initiate encoding and retrieval processes.

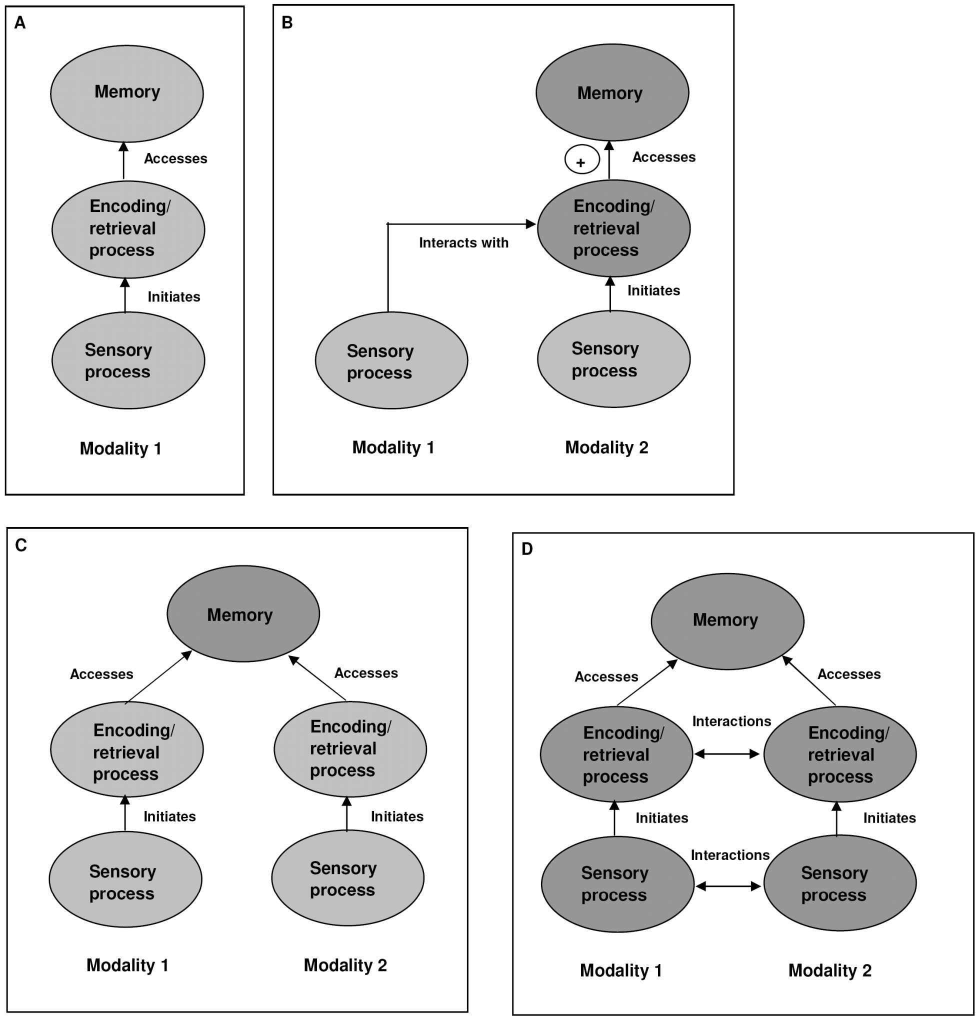


Figure 1.5. Mechanisms of uni- and multisensory encoding and retrieval. Panel A details the encoding/retrieval of unisensory information. *Sensory processing* of information initiates an *encoding/retrieval process*. *Memory* is viewed as a database containing information about the environment (i.e. object identities and locations). The *encoding process* adds information to this database. The *retrieval process* compares the information from sensory processing with information in memory. Successful retrieval allows an observer to recognize the objects and locations from the environment. When multiple sensory modalities process information two possible mechanisms may boost the effectiveness of the *encoding/retrieval process*. First, in panel B, the information perceived in the second modality may improve the encoding/retrieval of information initiated by the first modality (for instance, by setting additional constraints when searching information). Second, in panel C, each sensory modality that processes information from the environment may initiate its own *encoding/retrieval process*. Both mechanisms are not mutually exclusive. They are combined in panel D. In that panel, the *encoding/retrieval processes* initiated by *sensory processing* in each modality boost each other (in this case, memory performance may be improved because each process accesses a different part of the database). Also, in this panel, we added the *sensory interactions* which take place during sensory processing for completeness.

Estimate current position
The current position can be determined in two ways. First, *path integration* produces an estimate of the navigator's displacement. Second, we can recognize a location (e.g. from memory). Since we anticipate that multisensory presentation improves retrieval of information with unisensory presentation, we think that multisensory presentation can help a navigator to estimate his current position.

Target generation
One way or the other, a navigator determines where he wants to go. In TOM, this is handled by *target generation*. To set a *next target position*, *target generation* may need to access *memory*. For instance, if a navigator wants to go to the airport, he accesses the location of the airport in memory and sets that as his *next target position*.

Plan next movement
To plan his next movement, the navigator combines the estimate of his *current position* with the *next target position*. Additional *route information* may be involved in this process when, for example, there is no straight path from the *current position* to the *next target position.*
TOM indicates that multisensory presentation can help a navigator to plan a more suitable route because of two reasons. First, multisensory exploration of an environment may improve the quality and quantity of the information about the environment stored in *memory*. If more and better information about the environment is available, a better route can be selected. Second, finding a way through a multisensory instead of a unisensory environment may make the information in *memory* easier to retrieve. This also allows the selection of a better route. Both these effects are indirect results of multisensory presentation improving memory encoding and retrieval.

Move
After movement is planned, it is executed. Movement execution affects the navigators position and orientation in the environment. Because we expect that multisensory presentation leads to the selection of better routes than unisensory presentation, we expect that multisensory presentation allows a navigator to move more efficiently to the *next target position*, which should improve his navigation performance.

1.3 Critical research issues and outline of this thesis

The overarching purpose of this thesis is to investigate the benefits of congruent visual, auditory and tactile presentation on navigation in virtual environments. In the previous section, we introduced the Trisensory Orientation Model (TOM). TOM shows that the possible multisensory benefits in navigation rely on multisensory benefits in perception and memory. Multisensory benefits on perception and memory are therefore relevant to the purpose of this thesis.

The discussion that accompanied the introduction of TOM revealed several gaps in our knowledge about possible benefits of congruent presentation of visual, auditory and tactile information at the perceptual, memory, and navigation level. In the remaining chapters of this thesis we investigate seven research questions that are related to these gaps. Below, we give a short description of these chapters and formulate the research questions and accompanying hypotheses.

Perception

Most research on multisensory perception has focused on sensory interactions between vision and sound. Only a few studies have addressed interactions between vision and touch. Of particular interest is a study by Violentyev and colleagues (2005). They presented a visuotactile flash illusion and found that behavioral responses to that illusion matched those to a similar audiovisual flash illusion. This indicates that the neural mechanisms underlying audiovisual and visuotactile interactions may be similar. In chapter 2, we employed electroencephalography (EEG) measurements to study early visuotactile integration processes of the visuotactile flash illusion; we compare our results with those obtained in earlier EEG studies of the audiovisual flash illusion (i.e. Shams *et al.* 2001; Mishra *et al.* 2007). We studied these processes because early sensory interactions can play a role in subsequent memory and navigation processes and we wanted to know whether the results for audiovisual stimuli can be generalized to visuotactile stimuli.

Thusfar, multisensory perception has often been studied by presenting *incongruent* information to different sensory modalities (e.g. Ernst & Banks, 2002; Shams, Kamitani & Shimojo, but see Hershenson, 1962). In the real world, however, multisensory information usually originates from the same object or event and is usually *congruent*. Yet, little data is available specifically about the effect of congruent multisensory presentation. In chapter 3 and 4, we address this issue by studying the effects of congruent visual, auditory and/or tactile presentation in a temporal numerosity judgment task. The results we obtained in chapter 3 did not match those reported in a similar paradigm where incongruent stimuli were presented (i.e. the flash illusion). Therefore, we investigated whether different multisensory effects exist for incongruent and congruent multisensory presentation or whether these differences were due to other differences in the experimental settings.

The above paragraphs led to the following two research questions and hypotheses:

Question 1: Are the effects of visuotactile presentation reflected in early EEG patterns? We hypothesize that visuotactile interactions modulate activity along the visual cortex in a similar vein as audiovisual interactions, although the modulation may occur slightly later (about 10 ms), because tactile signals presented to the fingers take longer to reach the brain than visual or auditory signals.

Question 2: Does congruent multisensory presentation improve perceptual task performance? For temporal numerosity judgment, we hypothesize that an observer's ability to estimate numerosity improves when temporal series are congruently presented in multiple sensory modalities.

Memory

The known benefits of multisensory memory are limited to the effects of audiovisual presentations on encoding of object identity memory. In chapter 5, we presented visuotactile Morse codes to investigate whether these effects also apply to visuotactile presentation, to retrieval and to object location memory.

Previously, memory benefits were reported for semantic congruent multisensory cues, but not for combining semantic with non-semantic cues Laurenti *et al.* 2004; Lehmann & Murray, 2005, Yuvall-Greenberg & Deoull, 2007). Because for the latter, it is not clear whether the cues were congruent or not, we compare multisensory memory effects for congruent semantic and congruent non-semantic items in Chapter 6.

The following research questions and hypotheses were investigated:

Question 3: Do multisensory memory effects extend to visuotactile presentation?

We hypothesize that visuotactile presentation improves memory performance over visual and tactile presentation.

Question 4: Does multisensory presentation improve memory *retrieval* as well as encoding? Based upon the similarity between encoding and retrieval processes, we expect that multisensory memory effects are also present for retrieval.

Question 5: Does multisensory presentation improve object *location* memory as well as object identity memory? Since location information is subject to sensory interactions, we hypothesize that multisensory presentation can improve object location memory.

Question 6: Do multisensory memory benefits differ for congruent *semantic* and *non-semantic* items? In chapter 5 we find significant, but relatively small, multisensory benefits for congruent non-semantic Morse codes. We think that multisensory memory benefits for congruent semantic items are larger.

Navigation

Finally, at the navigation level, we have a single research question:

Question 7: Do multisensory landmarks increase navigation performance?

This question is investigated in chapter 7. Participants explored several virtual mazes which either contained visual, auditory, or audiovisual landmarks. Thereafter spatial memory and navigation performance was measured in four tasks. We hypothesized that navigation performance increases following exploration of a multisensory environment. According to TOM, multisensory presentation improves memory which allows a navigator to make better navigational decisions.

Chapter 2: Early sensory interactions of the touch-induced flash illusion: An EEG study

Abstract

Pairing two brief auditory beeps or tactile taps with a single flash can evoke the percept of a second, illusory, flash. Investigations of the neural mechanisms that underlie the influence of audition on vision have shown that auditory information can modulate activity in the visual cortex. In this chapter, we investigated whether touch modulates the visual evoked potential in a similar vein. Electroencephalogram (EEG) was recorded over occipital and parieto-occipital areas of 12 observers while they judged the number of flashes in stimuli consisting of tactile and/or visual pulses. We compared bisensory EEG to its unisensory constituents and found significant positive deflections around 110 ms and 200 ms and negative deflections around 330 ms and 390 ms from stimulus onset. Furthermore, comparison of the EEG activity in the trials in which the illusion was perceived and those that were physically the same but in which the illusion was absent revealed significant differences around 70 ms and 20 ms pre-stimulus onset and around 160 ms post-stimulus onset. Overall, the results suggest that touch can modulate activity along the visual cortex and that similar neural mechanisms underlie perception of the sound- and touch-induced flash illusion. The reported differences in pre-stimulus activity may reflect that the brain state just prior to stimulus onset affects the susceptibility to the flash illusion.

2.1. Introduction

Our brains provide an interpretation of our surroundings by integrating all aspects of the environment that the senses register into a coherent whole (Ernst & Bülthoff, 2004). Such sensory integration occurs in several midbrain structures such as the superior colliculus (Fries, 1984; Stein & Meredith, 1993) and higher-order association areas of the neocortex located in the parietal (Schroeder & Foxe, 2002), temporal (Schroeder *et al*., 2001; Schroeder & Foxe, 2002) and frontal lobes (Benevento, Fallon, Davis & Rezak, 1977).

Investigations of the (non-human) primate cortex indicate that multisensory integration can also occur in cortical areas that are regarded as unisensory. By means of single-cell and cell-cluster recordings, Fu and colleagues (2003) showed that neurons in the macaque auditory cortex also respond to touch. Sauvan and Peterhans (1999) recorded activity in cells of the prestriate cortex and found neurons that exhibit a selective responsiveness to specific stimulus orientations, irrespective of body position. This indicates a direct modulation by vestibular and/or proprioceptive signals of neurons in the monkey visual cortex. Using tracer injections, direct connections between primate auditory and visual cortex have also been identified (Falchier, Clavagnier, Barone & Kennedy, 2002; Rockland & Ojima, 2003).

In line with these primate findings, recent investigations of the human cortex indicate that human ‘unisensory’ brain areas are also affected by multisensory integration. (see Schroeder & Foxe, 2005; Ghazanfar & Schroeder, 2006; Macaluso, 2006). For instance, pairing visual (Schaefer, Flor, Heinze & Rotte, 2005) or auditory stimuli (Hötting, Friedrich & Röder 2009) with touch stimuli affects the event-related potential (ERP) in the somatosensory cortex.

Flash illusion paradigm

A paradigm used to reveal the processes by which (human) integration of visual and auditory information is accomplished is the sound-induced flash illusion (Shams, Kamitani & Shimojo, 2000; 2002). When people are presented with a single physical flash, presented together with two beeps presented in rapid succession (60-100 ms) they often report perceiving two flashes. Several brain mapping studies investigating this cross-sensory illusion report modulating effects of sound in the visual cortex (Arden, Wolfe & Messiter 2003; Bhattacharya, Shams & Shimojo, 2002; Mishra, Martinez, Sejnowski & Hillyard, 2007; Mishra, Martinez & Hillyard 2008; Shams, Kamitani, Thompson & Shimojo 2001; Shams, Iwaki, Chawla & Bhattacharya, 2004; Watkins, Shams, Tanaka, Haynes & Rees, 2006).

Of particular interest is the study by Shams *et al*. (2001), whom investigated whether this cross-sensory influence of audition on vision occurred at an early stage of processing or in later stages of processing. They recorded EEG over the visual cortex and compared the event-related potential (ERP) following the presentation of a single flash accompanied by two beeps with the ERP following the presentation of a single flash and no sound. Differences were found from 170 ms post-stimulus onward. In other studies (Bhattarcharya *et al*., 2002; Shams *et a*l., 2005), the ERP of a single flash and no sound was compared with the ERP following trials in which the flash illusion was actually perceived. Here, differences were found as early as 30-60 ms post stimulus onset. Differences at similar intervals were also reported by Mishra and colleagues (2007), whom compared the EEG associated with perceiving a flash illusion with the EEG when the perception of such an illusion is absent. All these studies suggest that the occurrence of the sound-induced flash illusion is associated with the post-stimulus activity in the visual cortex.

Touch-induced flash illusion

Most of the above studies reported on sensory interactions between audition and vision. Several behavioral and brain imaging studies show an influence of touch on vision (Bauer, Oostenveld & Fries, 2009; Philippi, Van Erp & Werkhoven, 2008; Sathian & Zangaladze, 2001; Van Erp and Werkhoven, 2004). A flash illusion can also be evoked by pairing a single flash with two tactile pulses. Behaviorally, the touch-induced flash illusion effect is similar to the sound-induced flash illusion (Bresciani, Dammeier & Ernst, 2006; Violentyev, Shams & Shimojo, 2005; Werkhoven, Van Erp & Philippi, 2009), but data on possible early interactions along the visual cortical areas are not yet available. In the first chapter of this thesis, we argued that interactions between vision and sound, and between vision and touch have the potential to improve 'visual' virtual environment applications. From this perspective it is relevant to know how the mechanisms underlying audiovisual interactions relate to those underlying visuotactile interactions.

Therefore, we presented participants, among other stimuli, with the touch-induced flash illusion and investigated the neural mechanisms underlying this tactile-visual illusion. Because tactile signals take longer than auditory signals to reach the brain we expect tactile-visual integration to occur later than auditory-visual integration. Based on the conduction speed of the nerves involved and the distance from the finger to the brain this difference may be in the order of 10 ms (i.e., Afferent axons have a speed of about 100 m/s; conduction time over a distance of 1m is 10 ms). This delay is also reflected in behavioral data: response time to a tactile signal is larger than for the visual and auditory channels (e.g., Goldstone, 1968; Van Erp & Verschoor, 2004). We suspect that the mechanisms underlying tactile-visual integration are otherwise similar to those underlying audio-visual integration. Therefore, we hypothesize that there are significant differences between the bimodal EEG and the sum of its unisensory constituents, and that there are significant differences in the EEG associated with or without seeing a flash illusion.

2.2 Experimental Procedures

Participants

Twelve participants (mean age 24.3 years, SD = 2.9; six females) took part in the experiment. All participants reported normal sense of touch and normal or corrected-to-normal vision. Before taking part in the study, they gave their written informed consent.

Setup

The experiment was conducted in a dimly lit, sound-attenuated, and electrically shielded chamber. Participants were seated approximately 52 cm in front of a monitor. Participants were instructed to maintain fixation on a white fixation cross, displayed at the center of a monitor against a black background. A white LED, positioned approximately 8° below fixation, was used to present the visual stimuli. These consisted of brief flashes of light, each with a duration of 10 ms and a luminance 80 cd/m². Tactile stimulation consisted of 10 ms (single 100 Hz sine-waves) taps on the tip of the right index finger, generated by a Mini-Shaker 4810 (Brüel and Kjær, Nærum, Denmark). Each tap had an amplitude of 0.06 mm and an acceleration of 25 m/s². Participants were provided with an arm rest to support their right arm. Any sound generated by the tactile actuator was masked by constant pink noise played through loudspeakers. In addition, participants wore foam ear plugs. Responses were given with the left hand on a numerical keypad.

Task and stimuli

Nine different stimuli were presented, consisting of all combinations of 0, 1, or 2 taps (T_0; T_1; T_2, suffixes denote number) and 0, 1, or 2 flashes (F_0; F_1; F_2). Multiple pulses in a single modality were separated by a 60 ms interval. In bisensory trials with an equal number of taps and flashes, pulses were presented simultaneously in both modalities. In trials where only one flash or tap was presented together with two taps or two flashes, the single flash or tap was presented in between the taps or flashes. Five hundred ms after stimulus offset, the white fixation cross was replaced by a question mark for 2 s during which the response was given.

The stimuli were presented in ten blocks of 135 trials. In each block, each stimulus was presented 15 times, resulting in 150 repetitions for each stimulus. Stimuli occurred in random order. The interval between stimulus presentations varied between 2310 and 2500 ms to counter any possible anticipation effects (Teder-Sälejärvi, MacDonald, Di Russo & Hillyard, 2002). Between blocks, participants had short breaks.

Participants received written instructions at the beginning of the experiment asking them to only report the number of flashes they perceived, and to ignore the taps. When no flashes were perceived, participants were instructed to press ‘0’. Before starting the experiment, participants were familiarized with the task by judging the numerosity of ten stimuli consisting of one or two flashes. This procedure was terminated once 90% accuracy was achieved. This took one or two blocks. Next, a block of 90 test-stimuli were presented to test their susceptibility to the flash illusions.

In order for volunteers to be allowed to participate we employed the criterion that two flashes had to be reported in at least 20% of the F_1T_2 flash illusion trials and that a single flash had to be reported in at least 20% of the F_2T_1 ‘suppressed’ flash illusion trials (see Andersen, Tiippana & Sams, 2004). This criterion was set to allow us to compare EEG activity in trials in which an illusion was reported to those in which no illusion was reported with sufficient statistical power. All participants met this criterion and participated in the experiment proper.

Once participants completed the test-trials, the electrodes were applied and the experiment proper started. The experiment lasted about 4 hours. One hour was required for instruction, familiarization with stimuli and testing flash illusion susceptibility, one hour for the application of the electrodes and about two hours for running the experiment.

Electrophysiological recordings

The electroencephalogram (EEG) was recorded from 3 occipital (O1, Oz, O2) and 3 parieto-occipital (PO3, POz, PO4) sites, according to the international 10-20 system. These sites cover the visual cortex, in which cross-modal effects have been observed before resulting from auditory stimulation (Shams, Kamitani, Thompson & Shimoko, 2001; Mishra, Martinez, Sejnowski & Hillyard, 2007). The signal was amplified using a g.USBamp (g.tec, Schiedlberg, Austria). Additional electrodes were placed below the outer canthus of the right eye and above the outer canthus of the left eye, to record eye-movements (EOG). Linked mastoids served as reference and the ground was positioned at Fp1. Electrode impedance was maintained below 5 kΩ.

The signal was digitized at 512 Hz, with a band-pass of 0.1-80 Hz. A 50 Hz notch filter was applied to eliminate main interferences.

ERP analysis

Trials containing amplitudes exceeding ±50 µV in EOG channels, or ±100 µV in any other channel were rejected. On average, 89.3 (SD=55.7) out of 1350 trials for a single observer were removed. Eye-blinks and/or artifacts were distributed evenly across stimulus types. Raw data were segmented into 3.5 s epochs (1.1 s pre-stimulus - 2.4 s post-stimulus), sorted according to type of stimulus and baseline-corrected (100-0 ms pre-stimulus). Because the unisensory trials were to be contrasted against the flash illusion trials, the F_1T_0 was recalibrated such that the flash occurred at the same time as the flash of the F_1T_2 stimulus. All trials were filtered for high-frequency noise using a moving average of five successive time samples. Finally, grand averages were computed for each stimulus, participant and electrode.

Three different analyses where employed to asses whether tactile stimulation modulated the activity along the visual cortex. These methods were frequently employed in brain-imaging studies of the sound-induced flash illusion and thus allow us to better discuss possible differences and similarities of the sound-induced and touch-induced flash illusions

Firstly, the grand averages of the unisensory components F_1T_0 and F_0T_2 were subtracted from the sum of the grand averages to the F_1T_2 and F_0T_0 'blank' stimuli (see Shams *et al*., 2001; Mishra *et al*. 2007). The result is referred to as the $[F_1T_2+F_0T_0]-[F_1T_0+F_0T_2]$ main difference wave. The grand averages of the F_0T_0 'blank' stimulus was included in the equation because an anticipatory ERP that precedes stimulus presentation would, when not added to F_1T_2-trials, be subtracted twice, possibly resulting in spurious cross-modal effects (Teder-Sälejärvi *et al*. 2002, Mishra *et al*., 2007; 2008).

Secondly, difference waves were also calculated for illusory flash trials (F1T2) in which only one flash was perceived (F_1T_2-R1) and in which two flashes were perceived (F_1T_2-R2) (Bhattarcharya, Shams & Shimojo, 2002; Shams, Iwaki, Chawla & Bhattarcharya, 2005).

These are referred to as the [F_1T_2-R1 +F_0T_0]-[F_1T_0+F_0T_2] and the [F_1T_2-R1 +F_0T_0]-[F_1T_0+F_0T_2] difference waves. Separate grand averages were calculated for the F_1T_2-R1 and F_1T_2-R2 trials of each participant and electrode.
Finally, the differences in activity in the visual cortex between either perceiving a flash illusion or not were also investigated by directly comparing the grand averages of the F_1T_2-R1 and F_1T_2-R2 trials (Mishra *et al*., 2007).

2.3 Results

Behavioral Analysis

Figure 2.1 displays the mean number of flashes reported after the presentation of each stimulus. Main effects of the presented Number of Taps, the presented Number of Flashes as well as their combined effect on the reported number of flashes were assessed by a Number of Taps (3) × Number of Flashes (3) repeated measures ANOVA. The main effects of Number of Taps and Number of Flashes were significant ($F_{2,\ 22.01} = 40.55$, $p < .001$, $F_{2,\ 22.02} = 1562.77$, $p < .001$, respectively), as was their interaction ($F_{4,\ 44.06} = 22.75$, $p < .001$).
A post-hoc Tukey HSD analysis revealed that the responses for all stimuli differed from each other ($p < .001$), except for the responses for the F_0T_0, F_0T_1, and F_0T_2 stimuli (all p's = 1.00); the F_1T_0 and F_1T_1 stimuli ($p = .15$); and the F_2T_0 and F_2T_2 stimuli ($p = .39$). In 40.3% of the presentations of the F_1T_2 stimuli, a touch-induced illusory flash was reported, which is less often than the 62.6% touch-induced illusions reported by Violentjev, Shams and Shimojo. (2005), but similar to the 37% and 34% reported for the sound-induced flash illusion by Mishra, Martinez, Sejnowski and Hillyard (2007), and Watkins, Shams, Tanaka, Haynes and Rees (2006). However, in contrast to Mishra *et al*., there was less variation in the proportion of illusions across participants: 58.3% instead of 34.4% reported an illusion in 30 to 50% of the flash illusions trials.

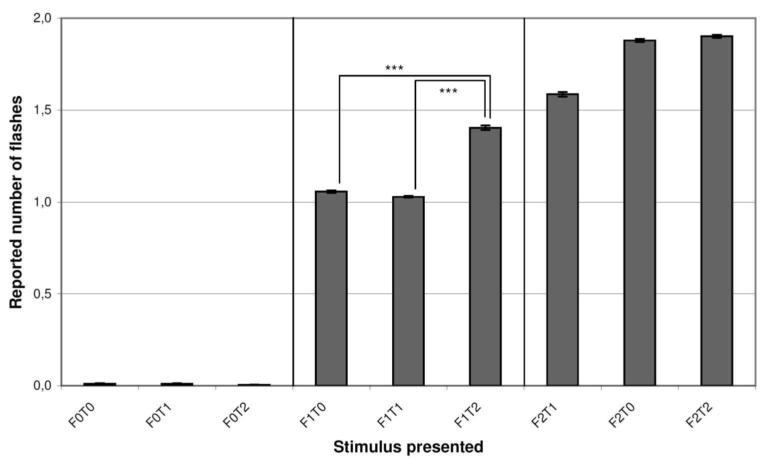


Figure 2.1: Reported number of flashes for each stimulus. Each stimulus is denoted by the number of flashes (F) and the number of Taps (T). Significant differences between the F_1T_0, the F_1T_1 and the F_1T_2 stimuli are indicated with asterisks. Three asterisks denote a p-level of <.001. Significant differences between other stimuli are reported in the text.

EEG difference wave analyses

Figure 2.2 displays the main, illusion, and no-illusion difference waves for the POz and Oz electrodes. In the main difference wave, the bisensory EEG of the F_1T_2 trial was compared with its unisensory constituents (F_1T_0 and F_0T_2) (see Shams, Kamitani, Thompson & Shimojo, 2001) and corrected for baseline activity with the EEG of a blank F_0T_0 trial (see Mishra *et al*., 2007). The illusion and no-illusion difference waves were similar to the main difference wave, except that only F_1T_2 trials were included in which a flash illusion was or was not reported (see Bhattarcharya, Shams & Shimojo, 2002 and Shams, Iwaki, Chawla & Bhattarcharya, 2005). Consequently, these two difference waves reflect the early bisensory activity along the visual cortex associated with or with not perceiving a flash illusion.

Single sample t-tests were used to determine whether the amplitude of the difference waves differed from zero at any time sample. Because applying Bonferroni-corrrections on very large sample sizes drastically compromises its statistical power we instead adopted the criterion for testing of difference potentials as proposed by Guthrie and Buchwald (1991). They ran computer simulations to determine how long an interval of consecutive significant points could be expected by chance. Based upon their results, we considered significant differences in 15 successive time samples as stable cross-modal interactions. Similar criteria have also been applied by other flash illusion studies (Shams *et al*., 2001; Bhattacharya *et al*., 2002; Shams *et al*., 2005).

Across electrodes and difference waves positive deflections (PD) were observed around 110 ms (PD110) and 200 ms (PD200) post stimulus onset. Negative deflections (ND) were observed around 330 ms (ND330) and from 390 ms post stimulus onwards (ND400). These cross-modal interactions are consistent with earlier reports on the sound-induced flash fusion illusion (Bhattacharya *et al*. 2002; Shams *et al*., 2001; 2005; Mishra *et al*. 2007).

The PD110 and the PD200 were mainly present in the main (PD200 only) and the illusion difference waves. Specifically, stable interactions were observed from 90 to 120 ms post stimulus onset in the illusion difference wave in the POz and PO3 electrodes; and from 150-230 ms post stimulus onset in the illusion difference wave (150-225 ms, all electrodes), the main difference wave (180-230 ms, all) and the no-illusion difference wave (200-230 ms, O1 and PO3 only).

In contrast, the ND330 and ND400 components were mainly present in the main and no-illusion difference waves. Specifically, stable interactions were observed from 310 to 360 ms post stimulus onset in the no-illusion (310-350 ms, all electrodes) and the main difference wave (310-360 ms; PO3, POz, PO4, O1, and O3); and from 375 ms post stimulus onwards in the no-illusion difference wave (375 ms onwards, all), the main difference wave (385 ms onwards, all), and the illusion difference wave (390 ms onwards, parietal only).

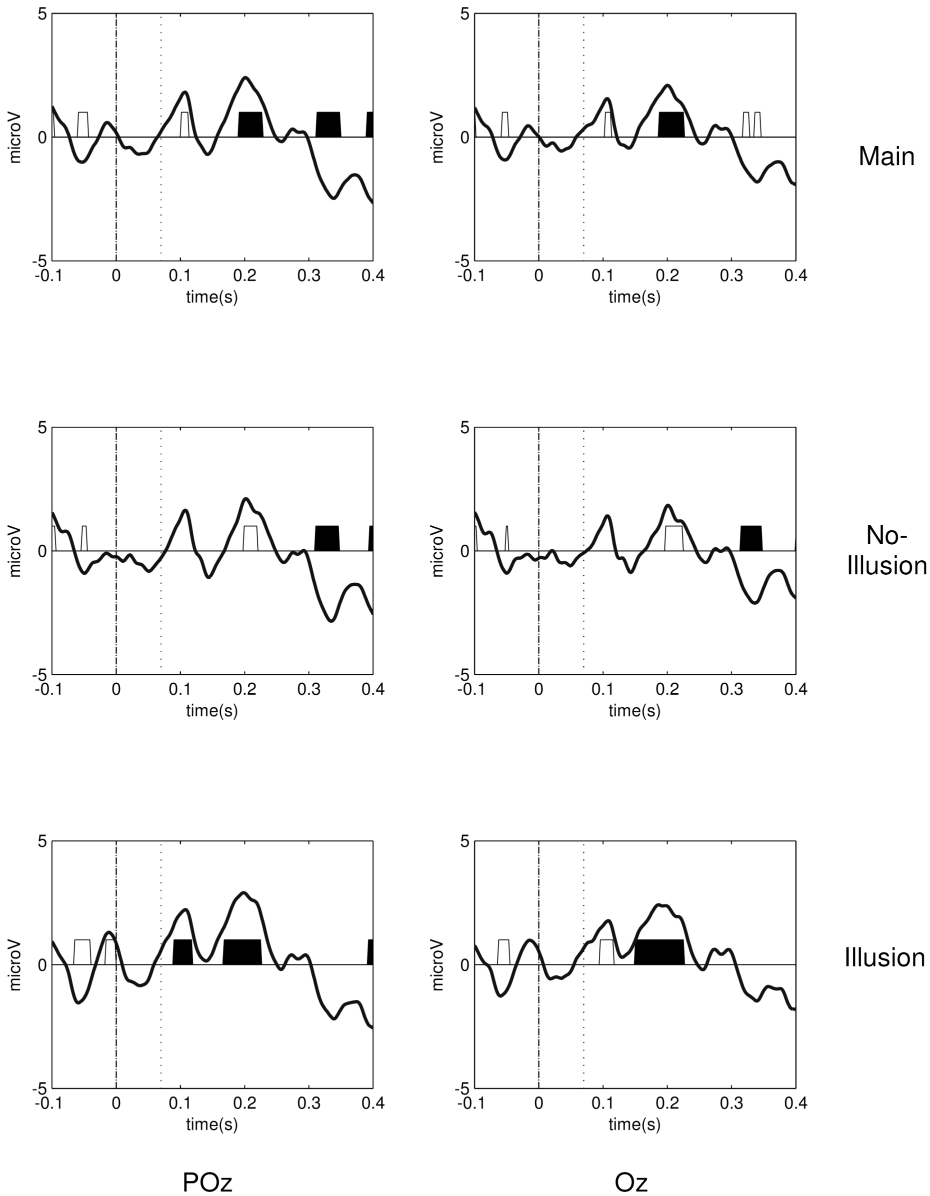


Figure 2.2: Various difference waves for the electrodes POz and Oz. The two upper panels show the main difference waves for the POz (left) and Oz (right) electrodes. The middle and lower panels show the no-illusion and illusion difference waves. The dot-dashed vertical line denotes stimulus onset while the dotted vertical line denotes offset. Black areas represent stable cross-modal interactions (i.e. significant differences [$p < .05$] for at least 15 successive time samples); white areas represent unstable cross-modal interactions (i.e. significant differences [$p < .05$] for less than 15 successive time samples).

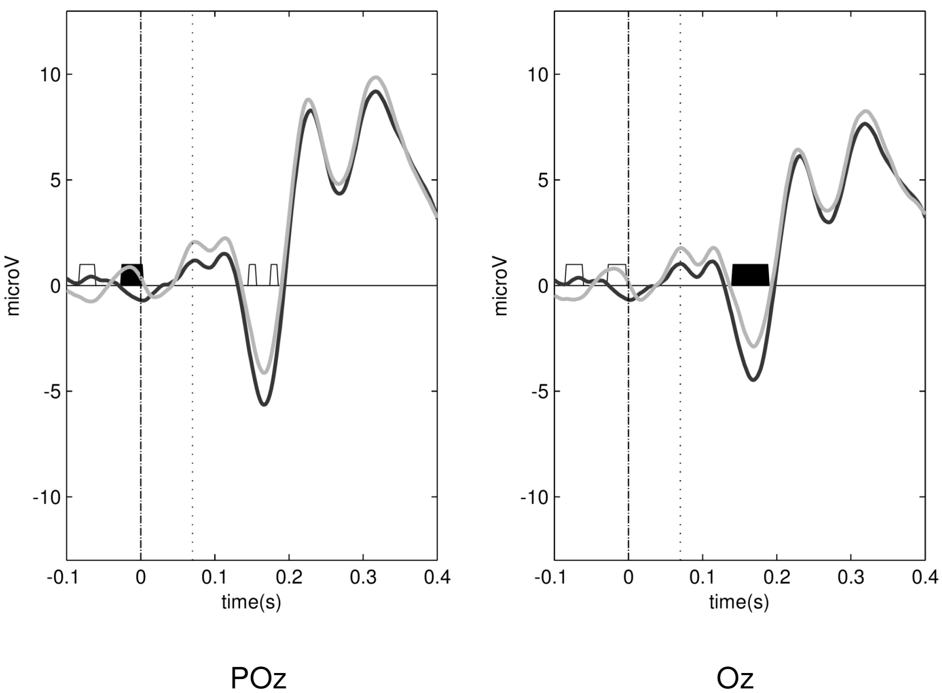


Figure 2.3: EEG activity associated with the perception of the second, illusory, flash (light gray line) and activity associated with no illusory perception (dark gray line) for the F_1T_2 stimulus. The dot-dashed vertical line denotes stimulus onset while the dotted vertical line denotes offset. Black areas represent stable cross-modal interactions (i.e., significant differences [p < .05] for at least 15 successive time samples); white areas represent unstable cross-modal interactions (i.e. significant differences [p < .05] for less than 15 successive time samples).

Trials with and without a reported flash illusion

In addition to the difference wave analyses, the perception of the flash illusion was assessed by directly comparing the ERPs for F_1T_2 trials in which a second flash was reported with ERPs for trials in which no second flash was reported (see Mishra *et al*., 2007). The EEG activity associated with each response for the POz and Oz electrodes is displayed in Figure 2.3.

Two sample t-tests were used to determine whether the response waves differed from each other at any time sample, again using the 15-sample criterion. We observed three different stable cross-sensory interactions. Stable interactions with larger negative amplitudes were observed from 140 to 185 ms (ND170) post stimulus onset in all occipital electrodes, but not in the parietal electrodes (only 3 to 11 successive samples). In addition, we found two stable interactions in two parietal electrodes before stimulus onset (-85 ms to -55 ms [D-70] and from -30 ms to 0 ms [D-15]). These two interactions were also observed in all other electrodes, but did not meet the criterion with only 9 to 14 successive samples.

2.4 Discussion

Pairing a single physical flash with two taps evoked perception of a flash illusion in 40% of the trials, which is in the usual range. In contrast with the recent study on the sound-induced flash illusion by Mishra *et al*. (2007), our individual observers varied less in the proportion that they perceived the flash illusion. It is possible that the testing procedure employed prior to the start of the experiment proper affected the probability of reporting a flash illusion.

We observed four cross-sensory interactions in the various difference waves and three additional interactions in the trial-based analyses. These indicate that touch alters the visual evoked potentials along the visual cortex. It is likely that some of these interactions occur in the visual cortex, even though the exact location cannot be determined due to the limited number of electrodes. Below, we discuss these interactions in more detail and compare them to the interactions reported in EEG studies for the sound-induced flash illusion.

Cross-sensory interactions in the difference waves

The earliest cross-modal interaction occurred at 90-120 ms post stimulus onset (PD110) in the illusion difference wave between the bimodal trials where flash illusions where reported (i.e., F_1T_2-R1 trials) and the unisensory components (i.e., the F_1T_0 and F_0T_2 trials). Thus, the interaction occurred as rapid as 40-70 ms after the second tap presented as part of the illusion inducing stimulus. Such an early interaction may be taken as evidence for a direct modulation by touch in the visual pathway (see Foxe & Simpson, 2002 on early interactions), as hypothesized by Violentyev, Shams & Shimojo (2005).The absence of this interaction in the main and no-illusion difference wave may indicate that this interaction is related to perceiving the flash illusion. A similar interaction was observed in the 'main' difference wave of the sound-induced flash illusion by Mishra *et al*. (2007), albeit it was observed slightly earlier after the presentation of the second beep (30-60 ms). Their results likewise suggested that this early interaction was related to the perception of the flash illusion.

A positive deflection was observed in the in the main difference wave at 185-225 ms (PD200). This interaction was also present in the illusion difference wave, but not in the no-illusion difference wave, which may suggest that the PD200 component may also be related to the perception of the flash illusion. Both Shams and colleagues (2001) and Mishra and colleagues observed similar deflections in the difference waves around 175 ms post stimulus onset. The results of Mishra and colleagues, however, indicated that this component of the sound-induced flash illusion is not associated with the perception of a flash illusion.

We also found interactions from 310-350 ms and from 390 ms post stimulus onwards onwards. Again, the results mirror previous findings concerning the sound-induced flash illusion (Shams *et al*. 2001, Mishra *et al*. 2007), although in the sound-induced flash illusion they were reported to occur earlier (starting from 270 ms post stimulus onset). Mishra *et al*. suggested that this interaction of the sound-induced flash illusion was unrelated to the perception of a flash illusion. In our data, these interactions occur mainly in the no-illusion and the main difference wave, which may actually indicate that they are related to the absence of the perception of a flash illusion.

Interactions between trials with or without a reported flash illusion

ERP analysis of trials in which a second flash was reported (F_1T_2-R2) compared to trials in which it was not reported (F_1T_2-R1), revealed interactions at 140 to 185 ms after stimulus onset. Importantly, these effects were mainly observed at the occipital electrodes, which may suggest interactions within the visual cortex. Trail-based results of the sound-induced flash illusion (see Mishra *et al.*, 2007), however, contrast ours as they reported earlier differences (90-150 ms) which originated from the auditory cortex and the superior temporal cortex. Although we did not measure activity from other areas than the visual cortex, the differences in timing may reflect specific differences between the sound- and touch-induced visual illusions.

We are the first to report pre-stimulus differences between the ERPs for trials in which a second, illusory, flash was reported and trials for which it was not reported. This difference suggests that susceptibility to the flash illusion may be related to the brain state prior to stimulus onset. Because the polarity of the activity associated with reporting the second flash differed from the polarity with not perceiving the flash illusion at the 85-60 ms pre-stimulus-interval and reversed at the 30-0 ms interval, these effects cannot be the result of the pre-stimulus baseline correction (-100-0 ms). Since these interactions were only stable for two electrodes it may be a highly localized effect. Because we recorded the EEG from only six sites we were unable to localize this effect reliably. Further studies using brain mapping tools with a high spatial resolution will be needed to investigate this effect.

Similarities and differences between the sound- and touch-induced flash illusions

While the overall effects associated with touch-induced flash illusion mirror those associated with the sound-induced flash illusion, comparisons revealed differences between the on- and offset of tactile-visual interactions and previously reported auditory-visual interactions associated with the perception of a flash illusion. These differences can partly be explained by the time required for auditory and tactile information to reach the visual cortex and are consistent with our hypothesis: tactile signals from the finger take longer to reach the brain than auditory signals and tactile-visual cross-modal interactions occur later than similar audiovisual interactions. However, considering the complexity of the pattern of cross-modal interactions of the flash illusion effect and the individual differences both to the susceptibility of the flash illusion (Shams *et al.*, 2000; 2002) those differences which cannot be explained by tactile delay will require a more detailed within-subjects analysis of the EEG associated with the sound- and touch-induced flash illusions.

Conclusion

In conclusion, we report early differences in the bimodal ERP (90-120 and 185-225 ms post stimulus onset) compared to the sum of its unisensory constituents, indicative of an early modulation of activity along the visual cortex by touch. These early difference are associated with perceiving a flash illusion. In contrast, the late differences (300 ms onwards post stimulus onset) we report may be related to the absence of perceiving a flash illusion. Furthermore, in comparison with the perception of just a single flash, the perception of the touch-induced flash illusion was associated with occipital, but not parietal differences which may suggest cross-modal interactions in the visual cortex. Although small differences exist, the present results mirror findings concerning the sound-induced flash illusion. The data thus supports the hypothesis that similar neural mechanisms underlie cross-modal integration of tactile and visual information as have been reported for integration of auditory and visual information. Our data further suggests a possible influence of brain state just before stimulus onset on the likelihood that a touch-induced flash illusion is reported.

Chapter 3. Multisensory temporal numerosity judgment[2]

Abstract

In temporal numerosity judgment, observers systematically underestimate the number of pulses. The strongest underestimations occur when stimuli are presented with a short interstimulus interval (ISI) and are stronger for vision than for audition and touch. We investigated if multisensory presentation leads to a reduction of underestimation. Participants were presented with 2 to 10 (combinations of) auditory beeps, tactile taps to the index finger and visual flashes at different ISIs (20 to 320 ms). For all presentation modes, we found underestimation, except for small number of pulses. A control experiment showed that the latter is due to a (cognitive) range effect. Averaged over conditions, the order of performance of sensory modalities is touch, audition and last vision. Generally, multisensory presentation improves performance over the unisensory presentations. For larger ISIs (160 and 320 ms) we found a tendency towards a reduction in variance for the multisensory presentation modes. For smaller ISIs (20 and 40 ms), we found a reduction in underestimation, but an increase in variance for the multisensory presentation modes. In the discussion, we relate these two findings to Maximum Likelihood Estimation (MLE) models which predict that multisensory integration reduces variance.

---

[2] Parts of this chapter have been published as:

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P.J. (2008). Multisensory temporal numerosity judgment. *Brain Research, 1242,* 116-126.

3.1. Introduction

In the 1940s, it became clear that observers had troubles adequately discriminating the short tones used in International Morse Code (Taubman, 1950a). Not only did this problem arise for the perception of auditory pulses, it also existed if the code consisted of flashes of light (called blinker code) (Taubman, 1950b). Common errors were incorrect responses to codes consisting of short pulses only, which were often perceived as codes consisting of fewer short pulses (Keller & Taubman, 1943; Keller & Schoenfeld, 1944). Additionally, it was reported that codes consisting of short pulses were sometimes overestimated with respect to the number of pulses. In other words, the number of pulses was under- as well as overestimated. Keller and Schoenfield termed these errors *dotting errors* and found that these made up a relatively large extent of all errors. Since training did not eliminate the dotting error (see Seashore & Kurtz, 1943), they suggested the replacement of several signals. This, however, never happened.

Temporal numerosity judgment

To investigate the cause of dotting errors, several studies in temporal numerosity judgment were carried out (Cheatham & White, 1952, 1954; Taubman, 1950a, 1950b). In these studies, a sequence of pulses (either flashes, beeps or taps) was presented and the observers' task was to estimate the number of pulses. Results suggested that there were significant differences in observers' performance between the three sensory modalities studied (see White & Cheatham, 1959). However, because experiments were carried out under strongly varying conditions, no direct comparison can be made.

Lechelt (1975) was the first to compare human performance in temporal numerosity judgment of visual, auditory and tactile stimuli under similar conditions. He presented two to nine pulses at rates of three to eight pulses per second. His findings can be summarized as follows:

1) The errors in estimation are in the direction of underestimation and the amount of underestimation increases as the number of pulses increases and/or as the interstimulus interval (ISI) decreases.
2) Perceived numerosity varies with presented numerosity and is unique for each modality.
3) The order of modality performance is consistent: audition is nearly perfect, followed by tactile and visual performance.

Modality differences in other tasks

Temporal numerosity judgment is not the only time related task for which modality differences have been reported. For instance, Van Erp and Werkhoven (2004) reported differences between touch and vision with respect to the estimation of empty time intervals. Modality differences in the estimation of empty intervals for audition and vision have also been reported (Behar & Bevan, 1961; Goldstone, Boardman & Lhamon, 1959; Goldstone & Goldfarb, 1963; Goldstone & Lhamon, 1972; Wearden, Edwards, Fakhri & Percival, 1998). Others have reported modality differences for duration discrimination (Droit-Volet, Meck

& Penny, 2007; Lhamon & Goldstone, 1974), temporal order judgment (Kanabus, Szelag, Rojek & Poppel, 2002), stimulus sequence identification (Garner & Gottwald, 1968; Handel & Buffardi, 1969), perception of temporal rhythms (Gault & Goodfellow 1938), and for temporal–tracking and continuation–tapping tasks (Kolers & Brewster, 1985). These studies show that in the time domain, best performance is reached with audition, followed by touch and then vision.

Sensory integration

Sensory systems complement each other as we naturally combine and integrate sensory information into a coherent and unambiguous percept (Ernst & Bülthoff, 2004). Recent studies in multisensory integration demonstrate that human perception can improve by measuring a single environmental property with more than one sense. Multiple (redundant) estimates available from multiple sensory modalities can shorten reaction time (Bernstein, Clark & Edelstein, 1969; Gielen, Schmidt & Van den Heuvel, 1983; Hershenson, 1962; Morell, 1968; Nickerson, 1973) or improve the reliability of the estimate (Alais & Burr 2004; Ernst & Banks, 2002; Gepstein & Banks 2003; Landy et al, 1995; Wu, Basdogen & Srinivasan, 1999). Multisensory stimuli can also enhance neural responses (Stein & Meredith, 1993) and enhance stimulus detectability (Vroomen & De Gelder, 2000) and perceived intensity (Stein, London, Wilkinson & Price, 1996).

As we mentioned in the first chapter of this thesis, *incongruent* stimuli were presented in many studies to explore multisensory interactions (Alais & Burr 2004; Ernst & Banks, 2002; Gepstein & Banks 2003; Landy et al, 1995; Wu, Basdogen & Srinivasan, 1999). Shams, Kamitani and Shimojo (2000, 2002), for instance, showed that we can perceive an illusory second flash when a single flash is presented together with multiple auditory pulses. However, many objects and events we perceive in our daily life provide congruent sensory information. Therefore, it is also important to study the effects of congruent multisensory presentation. Interestingly, the flash illusion experiments by Shams and colleagues can be viewed as a (limited) visual temporal numerosity judgment task with auditory *distracters,* which makes these studies and their explaining models relevant for our present numerosity judgment study.

Several Maximum Likelihood Estimation (MLE) models explain the flash illusion effect (Andersen, Tiippana & Sams, 2005; Ernst, 2006; Shams, Ma & Beierholm, 2005; Bresciani, Dammeier & Ernst, 2006). MLE models describe that the more reliable estimate (with the smallest variance) has a larger influence on the integrated percept. In the time domain, auditory estimates are generally more reliable than visual estimates giving them dominance over visual perception in the flash illusion experiments (Andersen *et al*., 2005). Similarly, the tactile modality is more reliable than the visual modality and can induce visual flash illusions (Violentyev, Shimojo & Shams, 2005). In turn, the tactile modality can be modulated by auditory stimuli (Bresciani *et al*., 2004). Interestingly, this suggests an order of dominance (in terms of influence) for numerosity estimates equal to the order of performance found by Lechelt (1975). That is, the more accurate modality in temporal numerosity judgment is the most influential in the flash illusion paradigm.

Multisensory numerosity judgment

Multisensory numerosity judgment has not yet been studied in the temporal domain and has only recently been studied in the spatial domain (Gallace, Tan & Spence, 2007). In *spatial* numerosity judgment, stimuli are presented simultaneously at multiple locations as opposed to sequentially at a single location as in *temporal* numerosity judgment. Gallace and colleagues were the first to report on multisensory spatial numerosity judgments. Participants were instructed to count the total number of pulses presented summed over the tactile and visual modalities (i.e., both modalities do not provide redundant information). They found that the amount of underestimation was larger for multisensory presentation than for unisensory presentation, which was hypothesized to be due to amodal/multisensory limitations in spatial processing (see Gallace *et al.*, 2007).

It is yet unknown to what extent congruent (i.e., redundant) multisensory presentation can improve temporal numerosity judgment. In the flash illusion paradigm, multisensory pulse sequences are often incongruent and participants are instructed to ignore the non-target modality. To investigate if and to what extent temporal numerosity judgment can benefit from multisensory presentation, we presented congruent multisensory stimuli and instructed participants that they should use this to their advantage. Stimuli consisted of 2 to 10 pulses with different ISIs under unisensory conditions (visual, auditory and tactile senses) as well as multisensory combinations. Based on general advantages found with multiple, redundant estimates, we expect congruent multisensory presentation to reduce the amount of underestimation as well as the variance in estimations over unisensory presentations.

3.1 Method

Participants

Twelve right-handed volunteers (seven male, five female) participated in the experiment. The participants were paid and naive as to the purpose of the experiment. None of the participants reported a history of sensory-motor disorders and all had normal or corrected-to-normal vision and normal audition and sense of touch. Participants had ages from 19 to 25 (mean 22.5) and gave their informed consent.

Setup

Participants were seated, in a dimly lit room, at a table, in front of a 15” computer monitor (Eizo, 640x480 pixels, 100 Hz frequency). The lower edge of the screen was 13 cm above the table and the monitor displayed a fixation cross in the center of the screen. A white LED was attached to the monitor eight degrees of visual angle below the fixation cross. A keyboard lay on the table in front of the monitor. When seated, the participant’s right arm was comfortably supported by an armrest while their right index finger (i.e., the index finger of their dominant hand) touched a tactile actuator (Bruel & Kjaer Mini-Shaker Type 4810). In order to mask the noise produced by the actuator, a speaker set produced constant pink noise at 47 dB(A) during all sessions. Participants wore a headphone (Sennheiser EH150) that provided the auditory pulses.

Stimuli

A custom-made computer program provided pulse series to the LED (displaying flashes), the headphones (producing beeps), the tactile actuator (producing taps), or a combination of them. Single pulses had a duration of 10 ms and, in the cases of beeps and taps, consisted of a single 100 Hz sine wave. Auditory, visual and tactile pulses were presented at intensities of 66 dB(A), 80 cd/m^2, and 12,5 m/s^2 (indenting the skin by 0,08 mm), respectively. All individual pulses were constant throughout the experiment and presented well above threshold. Unisensory pulse series consisted of 2 to 10 pulses (either flashes, taps or beeps) separated in time by an interval called interstimulus interval (ISI). Multisensory pulse series consisted of combinations of unisensory pulse series presented simultaneously with equal ISI and with an onset asynchrony of less than 1 ms. ISI was either 20, 40, 80, 160 or 320 ms. Figure 3.1 further illustrates the temporal intensity profiles of the stimuli.

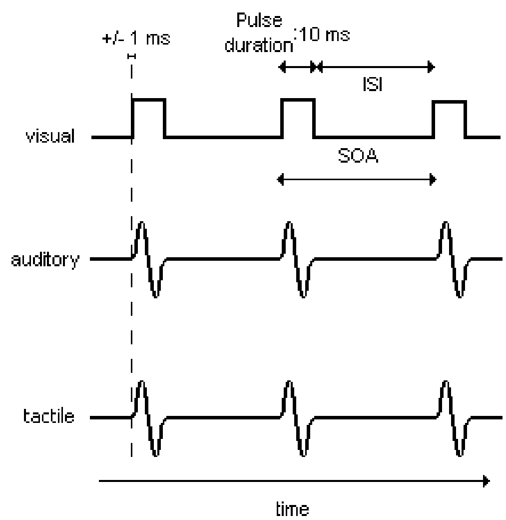


Figure 3.1 – Temporal profile of a multisensory stimulus. The amplitude of each signal is plotted on the y-axis. Each pulse had a duration of 10 ms. The ISI was varied over trials but was always equal for each modality. The example here corresponds to the presentation of three pulses in each modality (i.e., trisensory condition with three pulses)

Conditions

Participants were tested over seven sessions, corresponding to the seven modality conditions (three unisensory, three bisensory and one trisensory condition). In every session, each of the 45 pulse series (5 ISI x 9 N) was presented five times for a total of 225 stimulus presentations. The order of the pulse series in a session was random. At the start of each session and after each (small) break (see Procedures) an additional three random dummy stimuli were added, which were excluded from analysis. This was done to minimize the effects of observers accustoming to the task after each break or pause. The order of the sensory conditions (i.e., sessions) was semi-balanced across participants.

Procedures

At the start, participants had a short practice session of less than 10 minutes, allowing them to familiarize with the stimuli in all sensory conditions and the experimental setup. Participants were instructed to report the perceived number of pulses after each series via the numerical keypad on the keyboard and were told to press 0 when they perceived 10 pulses, effectively allowing them to report any number of pulses between 1 and 10. After each series was presented, a question mark replaced the fixation cross to indicate participants to give their response. Participants did not receive any feedback during the practice session or the experimental sessions. Before each session, they were told to which senses the pulses would be presented and were reminded to use all sensory information available. Each session lasted approximately 25 minutes and was divided in three blocks to allow participants for small breaks. More extensive breaks could be taken in between sessions.

Control experiment

A control experiment was conducted to investigate an unexpected overestimation of number of pulses presented in series with small ISIs and small numbers of pulses (see the results in section 3.3). Seven employees of TNO (6 female, 1 male, mean age 24.3) participated in the control experiment. The control experiment was similar the main experiment, but consisted only of two auditory conditions. These lasted about 25 minutes each and were taken on separate days. In one condition, only series consisting of 2 to 6 beeps were presented, while in the other, only series consisting of 6 to 10 beeps were presented. Each pulse series in each condition was presented 10 times at various ISIs (20, 40, 80, 160 and 320 ms).

3.3 Results

Pulse series were defined by the number of pulses presented, the type of presentation (i.e., the sensory modalities involved) and the ISI. Before analysis, we inspected the 18,900 responses for outliers. We computed group mean judgment and SD to each of the 45 unique series for each sensory condition. Two hundred sixty five responses (1.4%) deviated more than 3 SD from their group mean. These were considered as typing errors and were replaced with their respective group mean.

To test our hypotheses, we analyzed both mean and variance and interpreted the results with the following questions in mind:

1) Are there differences in bias and variance of numerosity estimation between sensory modalities (effects of individual modality characteristics)?
2) Does numerosity estimation for multisensory series differ from that for unisensory series (effects of multisensory integration)?

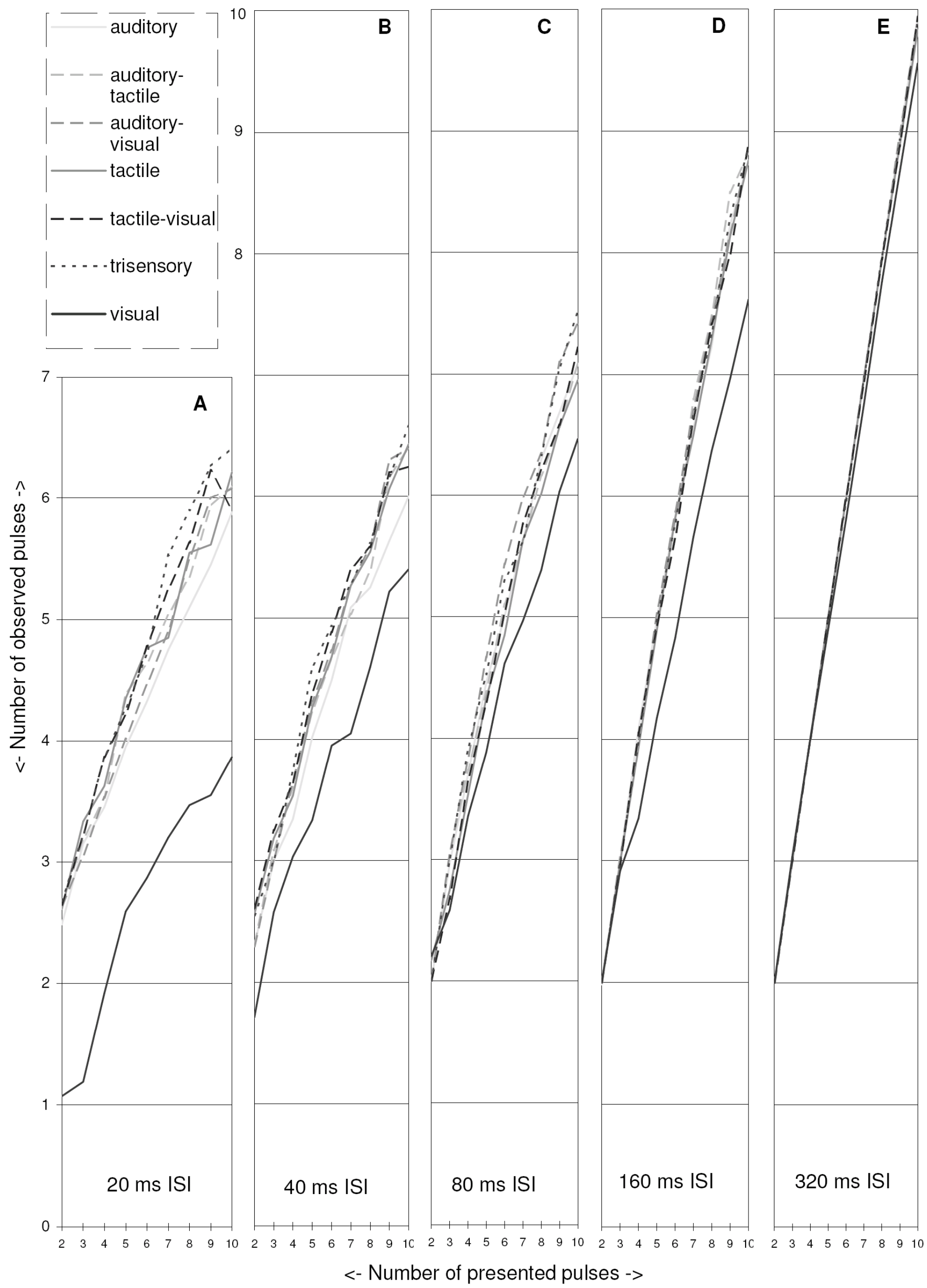


Figure 3.2 – Mean numerosity estimates. Each panel shows the number of pulses observed as a function of the number of pulses presented for a each ISI. In each panel, the responses for each sensory modality or combination of sensory modalities are connected with a grey scale-line.

Mean

Figure 3.2 shows the mean numerosity estimate for each sensory condition and each ISI as a function of the number of pulses presented. We observed a general tendency for underestimation, especially for smaller ISIs. To our surprise, we also observed overestimations for small numbers of pulses presented and for small ISIs. We hypothesized that participants became aware of the range of presentations, which might have influenced their judgment. To investigate this range effect we conducted a control experiment. The results of that experiment are reported at the end of this section.

The general relationship between the number of pulses presented (N) and the number estimated (E) given a certain ISI and modality combination can be well characterized as linear (Pearsons: all $p < .05$; r_{avg}=.82 r_{min}=.56; r_{max}=.99). The regression function calculated over all data is $E = 0.81 \times N$ (R=.773). We further analyzed mean estimates using an ANOVA with the following design: N (9) × ISI (5) × Sensory Condition (7). Significant differences were found for all effects (see Table 3.1). Post-hoc Tukey HSD analyses of the main effects revealed that every N and every ISI led to significantly different responses ($p < .001$).

Within the independent variable Sensory Condition, there were significant differences, which showed that:

1) the three unisensory conditions differed significantly ($p < .001$), and
2) the trisensory condition differed significantly from all its constituents ($p < .001$); the auditory-visual condition differed significantly from both its constituents ($p < .001$); the tactile-visual condition differed from both its tactile constituent ($p < .03$) and its visual constituent ($p < .001$) and the auditory-tactile condition differed significantly from the auditory condition ($p < .001$).

The above differences are depicted in Figure 3.3.

Table 3.1: N (9) x ISI (5) x Sensory Condition (7) ANOVA on mean

Factor	*df* effect	*df* error	F	*p* <
ISI	4	44	32.86	.001
Number of Pulses	8	88	542.06	.001
Sensory Condition	6	66	21.49	.001
ISI x N	32	352	62.58	.001
ISI x Sensory Condition	24	264	11.08	.001
N x Sensory Condition	48	528	3.29	.001
ISI x N x Sensory Condition	192	2112	1.64	.001

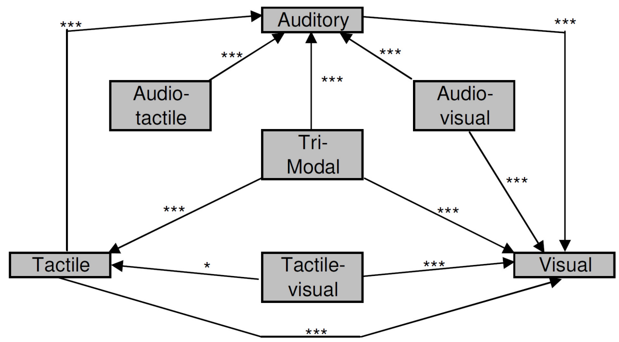


Figure 3.3 – Significant differences in mean between modality conditions. Only the unisensory differences and differences between unisensory and multisensory modality conditions have been included. If there are significant differences, the arrow points toward the modality condition with the largest amount of underestimation. One asterisk denotes a $p < .05$ level difference and three asterisks denote a $p < .001$ level difference.

Further visual inspection of Figure 3.2 suggested that the differences in the amount of underestimation between sensory conditions depended on the ISI. A post-hoc Tukey HSD analysis of the interaction Sensory Condition x ISI (see Table 3.2 for the mean responses to Sensory Condition x ISI) revealed the following significant differences:

1) With respect to the unisensory estimates, we found that the visual responses were significantly lower than the tactile and auditory responses (all $p < .001$), except for ISI 320 ms. Additionally, for ISIs of 20 and 40 ms, the tactile responses were significantly higher than the auditory responses.
2) With respect to differences between uni- and multisensory estimates, we found them to be significant (all at least $p < .01$) for ISIs of 20, 40, 80 and 160 ms (but not for 320 ms). These differences are summarized in Table 3.3.

Table 3.2: Mean and variance of all responses averaged over the number of pulses presented

	20 ms	40 ms	80 ms	160 ms	320 ms	All ISIs
Auditory	4.28 (.71)	4.35 (.53)	4.87 (.42)	5.63 (.25)	5.98 (.02)	5.02 (.38)
Tactile	4.55 (.59)	4.61 (.64)	4.76 (.50)	5.59 (.27)	5.96 (.04)	5.09 (.41)
Visual	2.63 (.44)	3.76 (.63)	4.39 (.61)	4.88 (.50)	5.83 (.16)	4.30 (.47)
Auditory-tactile	4.53 (.71)	4.54 (.64)	4.86 (.42)	5.70 (.18)	5.97 (.02)	5.12 (.39)
Auditory-visual	4.47 (.72)	4.62 (.62)	5.09 (.43)	5.68 (.21)	5.98 (.02)	5.17 (.41)
Tactile-visual	4.64 (.80)	4.69 (.71)	4.83 (.42)	5.62 (.23)	5.97 (.03)	5.15 (.44)
Trisensory	4.76 (.73)	4.72 (.60)	5.03 (.37)	5.66 (.18)	5.98 (.03)	5.23 (.38)
All Modalities	4.27 (.68)	4.47 (.63)	4.83 (.45)	5.53 (.26)	5.95 (.04)	5.01 (.41)

The first five columns display responses for five different ISIs. The last column gives the overall response for all ISIs. The rows display responses to each modality with the last row displaying the overall response for all modalities. The mean response per modality and ISI equals to the line average in Figure 3.2. The line average of an ideal observer is 6, the average of 2 to 10.

Table 3.3: An overview of the differences in mean between unisensory and multisensory conditions.

A	Auditory-tactile		B	Auditory-visual		
	Auditory	Tactile		Auditory	Visual	
20 ms	X		20 ms	X	X	
40 ms	X		40 ms	X	X	
80 ms			80 ms	X	X	
160 ms			160 ms		X	
320 ms			320 ms			
C	Tactile-visual		D	Trisensory		
	Tactile	Visual		Auditory	Tactile	Visual
20 ms		X	20 ms	X	X	X
40 ms		X	40 ms	X		X
80 ms		X	80 ms	X	X	X
160 ms		X	160 ms			X
320 ms			320 ms			

Each of the four panels show for each multisensory condition whether it differed significantly from it's unisensory constituents at certain ISIs. Note that if we found significant differences between multi- and unisensory responses, they were at the p < .01 level and the multisensory responses were always closer to the mean of an ideal observer than the unisensory responses (see also Table 3.2).

Variance

We analyzed the variance with a repeated measures ANOVA with the following design: N (9) x ISI (5) x Sensory Condition (7). Significant differences were found for both ISI and N, but not for Sensory Condition (see Table 3.4). Post-hoc Tukey HSD analyses for ISI revealed that all ISIs led to significantly different variances ($p < .01$), except for the ISIs 20 and 40 ms. Additionally, we found that the variance decreases as ISI increases. This is illustrated in panel A of Figure 3.4 (see also Table 3.2). Post-hoc analyses for N revealed significant differences when the number of pulses differed by four or more pulses ($p <.05$). Additionally, we found a consistently increasing variance as the number of pulses increased (see panel B in Figure 3.4).

Post-hoc analysis of ISI x Sensory Condition ($p < .001$) was conducted to investigate our two research questions. This analysis showed that:

1) With respect to unisensory estimates, there were no significant sensory differences within each ISI.
2) With respect to both uni- and multisensory estimates, there were only differences between the visual condition and the auditory-visual and tactile-visual conditions within the 20 ms ISI ($p < .05$).

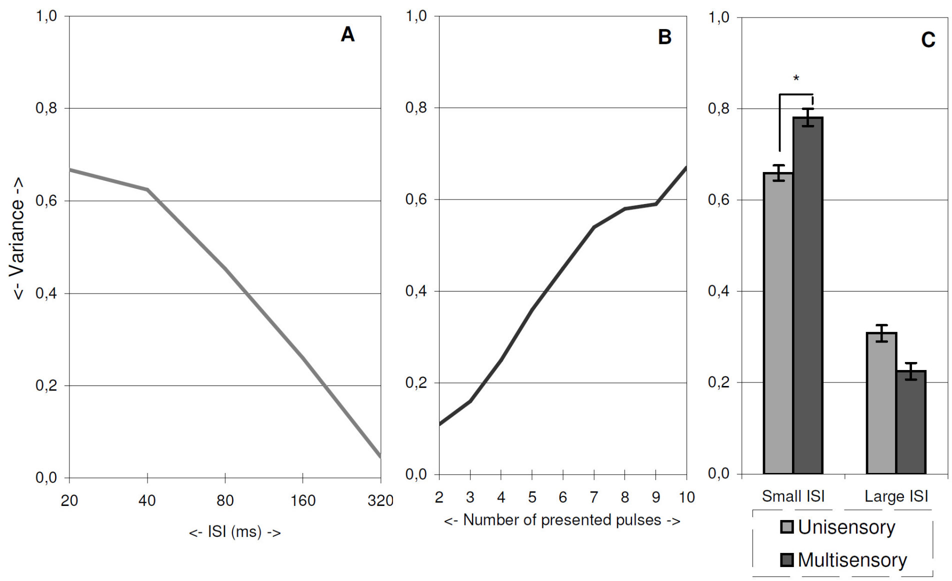


Figure 3.4 – Various variance relationships. For each of the three figures, the variance is plotted on the y-axis. Panel A displays the relation between the variance in responses and the ISI of the pulse series. Panel B displays the relation between the variance in response and the number of pulses in the pulse series. Panel C displays differences in variance in responses to unisensory and multisensory presentations at small ISI (20 and 40 ms) and large ISI (160 and 320 ms): For small ISI, the variance was larger for multisensory than for unisensory presentation, while for large ISI, this difference seemed reversed. The asterisk denotes a p <. 05 level difference.

Table 3.4: N (9) x ISI (5) x Sensory Condition (7) ANOVA on variance				
Factor	*df* effect	*df* error	F	*p* < or *p* =
ISI	4	44	39.00	.001
Number of Pulses	8	88	39.47	.001
Modality Condition	6	66	0.76	.605
ISI x N	32	352	3.27	.001
ISI x Modality Condition	24	264	2.63	.001
N x Modality Condition	48	528	0.86	.742
ISI x N x Modality Condition	192	2112	1.30	.005

In Table 3.2, however, we did notice a trend which was of interest to our research question: the multisensory variance was consistently smaller than the variance of their unisensory components for 160 and 320 ms ISIs, while for 20 and 40 ms ISIs the multisensory variance was consistently larger than the smallest unisensory component. For the 80 ms ISI, there was no such consistent trend. To investigate this trend we divided variance data into multisensory and unisensory, and in small ISIs (20 and 40 ms) and large ISIs (160 and 320 ms) and analyzed it with a Sensory Modality (2) x ISI (2) ANOVA. Here, the interaction Sensory Modality x ISI (p < .001) revealed that multisensory variance was significantly higher than the unisensory variance for the small ISIs ($p<.05$), while the multisensory variance showed a tendency to be lower than the unisensory variance for the large ISIs ($p < .1$). This is illustrated in Figure 3C.

Control experiment

An explanation for the overestimation found for small number of pulses is the range effect (resulting in regression to the mean of the scale). To investigate the presence of the range effect, we executed a control experiment. We presented participants with auditory stimuli in two different conditions at various ISIs (20, 40, 80, 160 and 320 ms). In one condition, the stimuli consisted of two to six beeps, while in the other condition the stimuli consisted of six to ten beeps. If participants would regress their response to the average number of presented pulses, we would find different responses to presentations of six beeps in both conditions.

Mean response to the presentation of six beeps was 5.28 in the low range condition and 5.65 in the high range condition. A Range (2) x ISI (5) ANOVA revealed that the main effects Range ($F_{1,6} = 10.14$; $p < .05$) and ISI ($F_{4,24} = 5.08$; $p < .001$) and the interaction ($F_{4,24} = 3.11$; $p < .05$) were significant. A post-hoc Tukey HSD test showed that there were significant differences due to Range for the ISIs of 20, 40 and 80 ms ($p < .001$). This is illustrated in Figure 3.5.

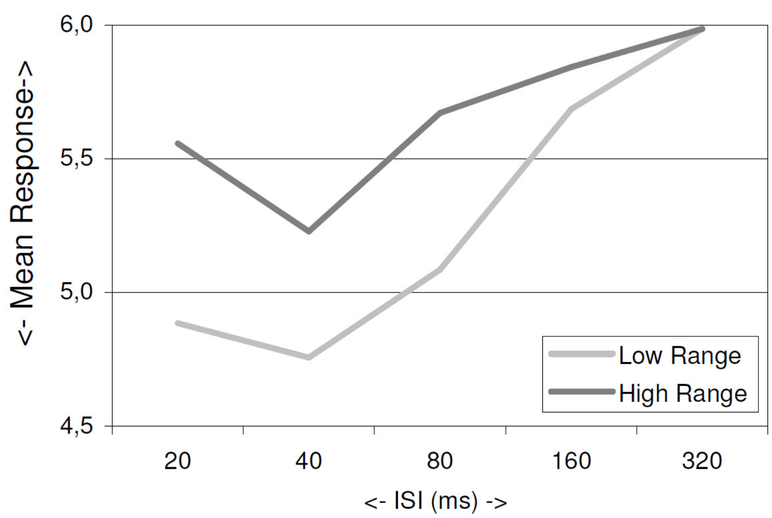


Figure 3.5 – Mean response to presentation of six beeps for each ISI. The dark gray line displays the results for the high range condition (which consisted of presentations of 6 to 10 beeps) and the light gray line displays them for the low range condition (which consisted of presentations of 2 to 6 beeps).

3.4 Discussion

General over- and underestimation

The data presented here confirm the general tendency towards underestimation for all (combinations of) sensory modalities, with the amount of underestimation increasing with the increase of the number of pulses and the decrease of ISI. This is in agreement with previous unisensory temporal numerosity judgment studies (Lechelt, 1975; White & Cheatham, 1959). However, we also found a small tendency towards overestimation for two to four pulses at small ISIs (20 and 40 ms). Overestimation of the number of multisensory pulses was also reported by Courtney, Motes and Hubbard (2007). They argue that differences in latencies of the different modalities cause the pulses perceived in different modalities to be (partially) added together. This, however, cannot explain our results because we not only find overestimation for multisensory, but also for unisensory presentations. The control experiment shows that the mean values at the lower and upper end of the range may have suffered from a range effect, resulting in over- and underestimation of the number of presented pulses, respectively. However, this range effect cannot explain the overall underestimation.

Only a few authors have attempted to find an explanation for the general underestimation. White and Cheatham (1959) attributed the similarity in underestimation of the major senses to ‘some temporal process’ in the central nervous system but made no attempt to actually explain the cause of underestimation. Efron (1970a) found that a single brief auditory stimulus evokes a perception with a minimum duration of 120-240 ms. For brief visual and vibrotactile stimuli slightly shorter durations (120-170 and 90-100 ms respectively) were obtained (Efron, 1970b; 1973; Gescheider, Valetutti, Padula & Verrillo, 1992). We hypothesize that this persistence of brief stimuli may hinder the clear separation between rapidly presented pulses. For ISIs that are much smaller than this persistence, pulses may even fuse together, ultimately resulting in a stream of individual pulses that are perceived as one continuous stimulus. This process resembles the flicker fusion threshold known in vision (De Lange, 1952; Levinson, 1968).

Differences in unisensory underestimation

The amount of underestimation differed per modality. We found that participants’ estimation of numerosity was better (i.e., less underestimation) for touch and audition than for vision when the pulses were presented at ISIs of 160ms or smaller. When pulses were presented at an ISI of 20 or 40 ms, the tactile estimations also had a smaller amount of underestimation than the auditory estimations. The order of performance for the major senses in our experiment thus is: touch, audition, vision.

This order of performance seems inconsistent with the order of performance reported by Lechelt (1975): audition, touch, vision. However, Lechelt tested temporal numerosity judgment for ranges from three to eight pulses per second which would be roughly equal to our ISI conditions of 160 and 320 ms. When we compare these data, they appear to be consistent: no differences between audition and touch. Furthermore, we are not the first to report the performance order touch, audition, vision. Earlier, Cheatham and White (1959) reported tactile data for counting taps at a rate of 30/sec (which is comparable to our ISI condition of 20 ms) which had less underestimation than their auditory data from a previous experiment at a similar rate (White & Cheatham, 1954). Alternatively, differences in the order of performance may relate to the location of tactile stimulation. In our experiment the taps were delivered to the index finger, while Lechelt stimulated the middle finger and Cheatham and White the thumb. Due to sensitivity differences, numerosity curves may depend on the finger stimulated.

Interestingly, we can relate the amount of underestimation in the different modalities and as function of ISI to the flicker fusion frequency and flutter threshold. For vision, the critical flicker fusion threshold lies around 50 Hz, but flicker fusion is still present for lower frequencies (De Lange, 1952; Levinson, 1968; Wells, Bernstein, Scott, Bennett & Mendelson, 2001). Our visual stimuli were presented at frequencies as high as 33 Hz and reduced flicker sensitivity might be the cause why the perception of visual stimuli tends to fuse much stronger than the perception of auditory or tactile stimuli. For the other two senses, a similar flutter threshold is believed to lie much higher (e.g., Gault & Goodfellow, 1938; Shipley, 1964).

The variance also increases with a decrease of the ISI, but also with an increase in the number of pulses. The latter relation is not only consistent, but also linear. Only for visual presentation and small ISI, we observe a small trend towards a decrease in variance with a decrease of ISI. Possibly, this is a result of the visual stimuli being close to the critical flicker fusion threshold. If a participant perceives one long pulse instead of many short ones, this should reduce the variance.

Differences between uni- and multisensory presentation

The major objective of this study was to investigate if multisensory presentation can increase temporal numerosity performance over unisensory presentation. The data show that it can. As the ISI decreases, the underestimation increases for all modality conditions, but more so for the unisensory conditions than for the multisensory conditions. We also found differences in variance, but these were dependant on ISI: For ISI smaller than 80 ms, multisensory variance increased in comparison to unisensory variance, while for ISI larger than 80 ms, we observe the opposite. Interestingly, only the data for large ISI are in accordance to the general predictions of the recently proposed MLE models (Andersen, Tiippana & Sams, 2005; Ernst, 2006; Shams, Ma & Beierholm, 2005; Bresciani, Dammeier & Ernst, 2006). These models predict that, if the noise is uncorrelated, each estimate has an influence on the integrated percept equal to their reliability. The integrated percept should therefore have a value that is in between the smallest and largest original estimate and the variance of the integrated estimate should be smaller than the variances of the original estimates. However, for small ISI, our data is not in accordance with these predictions. A possible cause may be that the MLE models assume unbiased estimators while the responses in our experiment were increasingly biased (i.e., more underestimation) as ISI decreased. In addition, the control experiment shows that responses may be influenced by the range effect to some extent. The range effect is a cognitive factor while MLE models have been tested only for perceptual phenomena.

Even so, appointing the influence of each modality is not trivial in our experiment. Suppose that we observe the following:

1) when one is presented with eight flashes one observes three,
2) when one is presented with eight beeps one observes five and
3) when one is presented with eight flashes and eight beeps one observes six.

Now to what extent is the observation ‘six’ influenced by the perception of flashes and the perception of beeps? As the multisensory observation is closer to the auditory observation one could claim that the beeps have a larger influence, but the influence is impossible to quantify. To further investigate this issue, future research might be required to look closer to the perception of the individual pulses than simply the number of pulses perceived. Which pulses are actually perceived and which pulses not?

EEG and fMRI studies suggest that the flash illusion effect is a result of dynamic interplay between auditory, visual and polymodal areas (Shams, Kamitani, Thompson & Shimojo, 2001; Mishra, Martinez, Sejnowski & Hillyard, 2007; Watkins *et al*., 2006). The correlation between conscious perception and (early) sensory activity in flash illusion experiments suggests that employing similar (EEG) techniques for temporal numerosity judgment could possibly answer the questions above. Interestingly, Noesselt *et al*. (2007) found that congruent audiovisual temporal streams of nonsemantic stimuli elicit increased activity in the multisensory superior temporal sulcus (STS) and primary auditory and visual cortices in comparison with unisensory streams. This, as opposed to incongruent audiovisual streams, which cause a decrease in comparison with unisensory streams. Since in our experiment multisensory presentation was congruent, it is possible that such increased activity in the primary cortices is related to the reduction of underestimation.

As for an explanation for the reduction in underestimation in multisensory presentation, it might be a possibility that a multisensory mechanism reduces or prevents the earlier mentioned fusion of pulses. Recently, Courtney, Motes and Hubbard (2007) showed that a continuous flash (317 ms) accompanied by two brief beeps (7 ms) can be perceived as two flashes. That is, participants experienced a ‘break-up’ of the continuous flash. If a continuous stimulus can be broken up, it is likely that two pulses which are perceived as one can be broken up in a similar way. Essentially, such an explanation would be an updated version of the discontinuity hypothesis as proposed by Shams *et al*. (2000, 2002), which stated that the discontinuous stimulus has a stronger effect on the final percept than the continuous stimulus. Because each modality is subject to fusion to some extent when presented with a pulse series, each modality could potentially influence every other modality. If two modalities simultaneously influence each other this could explain the reduction in underestimation.

Conclusion

Congruent multisensory presentations in temporal numerosity judgment can result in a reduction in variance (for large ISIs) or a reduction in underestimation (for small ISIs) as compared to unisensory presentations. Henceforth, we conclude that congruent multisensory presentation improves temporal numerosity judgment.

Chapter 4. Multisensory effects differ for counting small and large pulse numbers[3]

Abstract

In Flash Illusion (FI) experiments, congruent multisensory presentation has no effect on the mean estimate of the number of events, but decreases the variance in comparison with unisensory presentation. In the previous chapter, however, we found that congruent multisensory presentation in another Temporal Numerosity Judgment (TNJ) task can also affect the mean estimate (i.e., it reduced underestimation) and increases the variance. In this chapter we conducted three experiments to investigate the differences between both paradigms as possible causes of this discrepancy. These differences were: the presence or absence of incongruent stimuli (Experiment 4.1), the instruction to the observer to either count flashes, beeps, or multisensory events (Experiment 4.2), and the range of pulses presented (Experiment 4.3). We found significant differences between the mean numerosity estimate of multisensory and unisensory series in Experiment 4.3, but not in 3.1 and 3.2. This suggests that the difference in the range of pulses presented in FI (1-3 pulses) and TNJ (1-10 pulses) is the primary cause of the discrepancy. In the discussion we propose that this result may be explained by the use of two different strategies and their susceptibility to multisensory presentation. For small pulse numbers, observers can accurately count both unisensory and multisensory pulses. For larger numbers, observers can no longer count but will estimate the number based on the pulse series duration which is improved for multisensory stimuli.

---

[3] Parts of this chapter have been accepted as:

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. (*in press*). Multisensory effects differ for counting small and large pulse numbers. *Seeing and Perceiving*.

4.1 Introduction

We constantly perceive information about the world around us with a variety of senses and integrate this sensory information into a holistic view of the world (Ernst and Bülthoff, 2004). The mechanisms of multisensory integration can affect behavior. For example, it can shorten reaction time (Bernstein, Clark & Edelstein, 1969; Hershenson, 1962; Nickerson, 1973), enhance stimulus detection (Vroomen & De Gelder, 2000) and improve the reliability of a combined estimate, such as object size (Alais & Burr, 2004; Ernst & Banks, 2002; Gepshtein & Banks, 2003).

For counting pulses in a rapidly presented series, different multisensory effects have been reported in two different paradigms, viz. Flash Illusion (FI) and Temporal Numerosity Judgment (TNJ). In the FI paradigm, participants count the number of events in a target modality while ignoring the events in a distracter modality (Andersen, Tiipanna & Sams, 2004; Philippi, Van Erp & Werkhoven, 2010; Shams, Kitakami & Shimojo, 2000; 2002; Werkhoven, Van Erp & Philippi, 2009; Violentyev, Shimojo & Shams, 2005). Trials can either be congruent or incongruent (i.e., the number of events in the target and distracter modality is the same or different, respectively). In the incongruent trials, distracter events bias the estimate of the number of target events, while in the congruent trials, distracter events do not affect the mean estimate compared to the presentation of the target alone. In addition, in a single study it was reported that the distracter also increased the reliability (i.e., a reduction in variance) of the estimates in the congruent trials (Bresciani, Dammeier & Ernst, 2006).

In the TNJ paradigm, participants count the number of pulses in unisensory or in congruent multisensory pulse series (Cheatham & White, 1952, 1954; Lechelt, 1965; Philippi, Van Erp & Werkhoven, 2008). In the previous chapter we showed for series with Inter Stimulus Intervals (ISIs) similar to the FI paradigm (i.e., ISIs below 100 ms), that congruent multisensory presentation affected the mean estimate (i.e., the estimate of the number of pulses was higher in multisensory than in unisensory series). Furthermore, in some (but not all) of the comparisons between unisensory and multisensory trials we found a decrease of reliability (i.e., an increase in variance) for multisensory presentation.

Thus, the effects of multisensory presentation on the mean and variance of the numerosity estimates differ for the FI and TNJ paradigms as summarized in Table 4.1. In the present chapter, we investigate which of the differences between the FI and TNJ paradigm cause this discrepancy. There are three differences between the FI and the TNJ paradigms:

1) The presence or absence of incongruent stimuli: in the TNJ paradigm all stimuli consist of congruent pulse series, while in the FI experiments congruent as well as incongruent stimuli are presented within a session.
2) The instructions to the observer: in FI, participants are instructed to report the target modality only, sometimes accompanied by the explicit instruction to ignore the irrelevant or distracter modality. In TNJ experiments, participants are instructed to report the number of pulses with no reference to a sensory modality.

Table 4.1: Effects of rapid multisensory presentation (< 100 ms ISI) on mean estimate and variance for the Temporal Numerosity Judgment (TNJ) and Flash Illusion (FI) paradigms

Paradigm	Paradigm Characteristics	Mean estimate	Variance
TNJ	• congruent stimuli only • instruction to report the number of pulses • 1 to 10 pulses	Increases	Remains equal or increases
FI	• congruent and incongruent stimuli mixed • instruction to report the number of pulses in the target modality and ignore the irrelevant modality • 1 to 3 pulses	Remains equal	Remains equal or decreases

3) The range of the number of pulses presented: FI experiments are typically conducted with pulse series with a relatively small number of pulses (usually 1 to 3) compared to those in TNJ experiments (1 up to 10 or even 20 pulses).

Experiment 4.1

We hypothesized that each of these three differences could be the cause of the discrepancy and investigated these by conducting three experiments. In Experiment 4.1, we tested the hypothesis that the presence or absence of incongruent stimuli caused or contributed to the discrepancy. We presented one to three pulses in an auditory (1-A), a visual (1-V) and two audiovisual conditions. In one of the audiovisual conditions (1-AV-congruent) we only presented congruent stimuli, while in the other (1-AV-incongruent) we presented both incongruent and congruent stimuli within a session. Perhaps participants are able to detect the incongruence which could make them (more) eager to count only one modality and ignore the other (see Werkhoven *et al.*, 2009 on top-down effects on sensory integration). Their estimates of the congruent stimuli in the incongruent condition would then deviate less from their unisensory estimates than their estimates of the congruent stimuli in the congruent condition.

Experiment 4.2

In Experiment 4.2, we tested the hypothesis that the instruction to the observer caused or contributed to the discrepancy. We presented three audiovisual conditions with identical stimuli. Dependent on the experimental condition, we instructed participants to either count the flashes (2-AV-count flashes), the beeps (2-AV-count beeps) or to give a single holistic judgment about the number of events (2-AV-count events). In the latter instruction, which was similar to the instruction in the TNJ study by Philippi *et al.* (2008), we explicitly told participants that the number of beeps and flashes was always equal. The instruction to count only a single modality and to ignore the other encourages participants to suppress sensory integration (Werkhoven *et al.*, 2009). In contrast, the holistic instruction may actually encourage participants to integrate sensory information. If so, multisensory numerosity estimates will deviate less from the unisensory estimates when participants are instructed to count one modality and ignore the other than when they receive the holistic instruction.

Experiment 4.3

In Experiment 4.3, we tested the hypothesis that the difference in the range of the number of pulses caused or contributed to the discrepancy. We presented 1 to 10 pulses in an auditory (3-A), a visual (3-V) and an audiovisual (3-AV) condition and compared the results to the unisensory and congruent multisensory data from experiment 4.1 with its pulse range 1-3. In the previous chapter we hypothesized that multisensory presentation in TNJ affected the mean estimate by decreasing the amount of underestimation. Since the absolute amount of underestimation increases with the number of pulses presented there should be a difference in the effect of multisensory presentation on small and large numbers of pulses. That is, the (favorable) effect of multisensory presentation should be larger if the number of pulses is larger. If so, then the multisensory effects would differ for the large and small range because the large range contains more stimuli with larger numbers of pulses than the small range. To further explore this hypothetical mechanism we also investigated the results for the range 1-3 pulses of the three conditions in Experiment 4.3. If the multisensory effects depend on the number of pulses presented we should not find differences in the multisensory effects in the range 1-3 of experiments 4.1 and 4.3.

4.2 General Methods

Participants

Twelve right-handed participants (mean age = 23.7; SD = 2.25; 4 females) participated in all three experiments, spread out over at least nine days. The participants were naive to the purpose of the experiment and gave their informed consent after receiving instructions. All participants had normal or corrected-to-normal vision and normal hearing.

Setup

Participants were seated in a dimly lit room, in front of a 15” computer monitor (Eizo, 640x480 pixels, 100 Hz frequency) and a keyboard. The screen was not used for stimulus presentation, and only displayed a fixation cross. Eight degrees of visual angle below this cross a white LED was attached to the monitor. Participants wore headphones (Sennheiser EH150), during all three experiments.

Stimuli

Visual stimuli consisted of a series of flashes (or visual pulses) presented via the LED and auditory stimuli consisted of a series of beeps (or auditory pulses) presented via the headphones. Multisensory stimuli consisted of a series of flashes and a series of beeps. All series consisted of 1 to 10 pulses, which each had a duration of 10 ms and were separated by an ISI of 60 ms. Timing accuracy was better than 1 ms. Visual and auditory series in a multisensory stimulus were presented synchronously for equal numbers of pulses, with an onset asynchrony of 35 ms for incongruences of 1 pulse, or with an onset asynchrony of 70 ms for incongruences of 2 pulses (see Fig. 4.1). Individual flashes (80 cd/m^2) and beeps (67 dB(A), 100 Hz) were both presented well above threshold.

Conditions

In Experiment 4.1, four experimental conditions were tested: an auditory (1-A), a visual (1-V) and two audiovisual conditions. In the auditory condition participants were presented with series of 1, 2, or 3 beeps and were instructed to report their number. Likewise, participants had to report the number of flashes in series consisting of 1, 2, or 3 flashes in the visual condition. In both the audiovisual conditions participants were instructed to report the number of flashes and to ignore the beeps. In the incongruent condition (1-AV-incongruent) stimuli consisting of all combinations of 1, 2, or 3 flashes and 1, 2, or 3 beeps were presented, while in the congruent condition (1-AV-congruent) only stimuli with an equal number of flashes and beeps were presented.

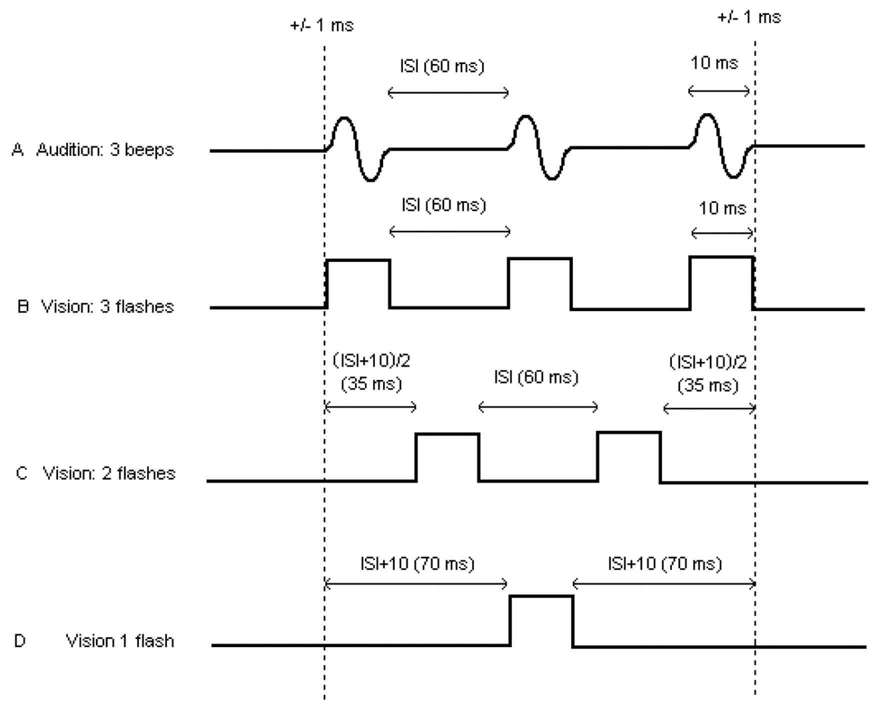


Figure 4.1 – Onset asynchrony in multisensory stimuli. Four series consisting of one to three flashes or beeps are displayed to illustrate the onset asynchronies in multisensory presentation for two series with an equal number of pulses (series A and B), or for series where the number of pulses differed by one (series A and C) or two (series A and D). Individual beeps are represented by a sine wave, while flashes are represented by a block wave. Stimulus durations and ISI are not to scale.

In Experiment 4.2, three audiovisual conditions were tested. In all conditions the stimuli consisted of those combinations of 1, 2, or 3 flashes and 1, 2, or 3 beeps such that the difference between the number of beeps and flashes was either 0 or 1. Dependent on the condition, participants were instructed to either count the number of flashes (2-AV-count flashes), the number of beeps (2-AV-count beeps), or the number of events (2-AV-count events). In the latter condition, they were specifically told that the number of flashes and beeps was always equal; a flash was always presented simultaneously with a beep and participants could consequently consider such a presentation as a single event. Interestingly, after this condition was conducted, almost none of the participants reported that they noticed that the number of flashes and beeps differed in that condition. In contrast, all of the participants reported that the number of flashes and beeps did differ in the other two conditions of Experiment 2.

In Experiment 4.3, three experimental conditions were tested: an auditory (3-A), a visual (3-V) and an audiovisual condition (3-AV). Unisensory stimuli consisted of series of 1 to 10 flashes or beeps and multisensory stimuli consisted of a series of 1 to 10 flashes and beeps (i.e., the number of flashes and beeps presented was always equal). Participants were instructed to count either the flashes (3-V and 3-AV) or the beeps (3-A).

Procedure

The three experiments were conducted in five sessions (see Table 4.2 for details). The sessions were separated by an interval of one to three days and each session took about 7 to 30 minutes to complete. The order of sessions and the order of conditions within the sessions were balanced across participants. The stimuli presented in each condition were each presented 10 times and were presented in random order. In addition, the first 10 stimuli of each condition were random duplicates and were not included in the analyses. These were included to allow participants to familiarize with the stimuli. Additionally, participants were presented with a small warming-up session before they started the experiment. The warming-up consisted of a visual phase and auditory phase whose order was balanced across participants. In each of these, participants were presented with 10 unisensory series of 1 to 3 pulses.

Before the start of the first session, participants received written instruction which further detailed the experiment. At the start of each condition, participants received a specific written instruction for that condition. They were instructed to either count the beeps, the flashes, or the events and to report the number perceived with their right hand via the numeric keypad on the keyboard. In all conditions, participants were told to press the 0-key when they observed 10 flashes, beeps, or events. After each stimulus presentation, the fixation cross was replaced by a question mark to indicate that participants had to report their estimate within 5 seconds. Participants were also informed whether the upcoming condition was unisensory or multisensory and were reminded to focus on the fixation cross.

Table 4.2: Details of the sessions and conditions of the three experiments.			
Session	Time to Complete	Conditions	Number of stimuli in condition
Training	2 minutes	Warming-up	20
Session A	10 minutes	Exp 4.1: Audition	(10 +) 30
		Exp 4.1: Vision	(10 +) 30
		Exp 4.1: AV congruent	(10 +) 30
Session B	15 minutes	Exp 4.1: AV incongruent	(10 +) 90
		Exp 4.2: AV count flashes	(10 +) 70
Session C	7 minutes	Exp 4.2: AV count events	(10 +) 70
Session D	7 minutes	Exp 4.2: AV count beeps	(10 +) 70
Session E	30 minutes	Exp 4.3: Audition	(10 +) 100
		Exp 4.3: Vision	(10 +) 100
		Exp 4.3: AV	(10 +) 100

Each row specifies the details of a session. The second column details the approximate time it took to complete it, the third column details which conditions were included in which sessions and the last column details the number of stimuli presented in each condition. Each session was run on a different day with 1-3 days in between sessions.

4.3 Results

Basic statistics

Two participants were removed from analyses because they did not consistently follow the instructions to complete the multi-day experiments. Table 4.3 displays the basic statistics for the congruent data. Note that the variance in the responses was first computed for each stimulus and each participant and than averaged over stimuli and participants (see Bresciani, Dammeier & Ernst, 2006). The mean responses listed in Table 4.3 indicate that that observers underestimated the presented number of pulses in Experiment 4.3 ($t_{(12)} = 2.7$; $p < .05$), but not in experiments 4.1 ($t_{(12)} = 0.1$; $p = .94$) or 4.2 ($t_{(12)} = 0.0$; $p = .97$).

We assessed the three hypotheses using repeated measures ANOVAs, with Tukey HSD post-hoc analyses when appropriate.

Table 4.3: Basic Statistics for the Conditions of the Three Experiments.					
Exp.	Condition	True Mean	Mean response (SD)	Variance (SD)	Accuracy
1	1-V	2	1.96 (0.85)	0.16 (0.15)	80 %
1	1-A	2	2.01 (0.84)	0.05 (0.07)	95 %
1	1-AV-congruent	2	2.01 (0.85)	0.10 (0.11)	86 %
1	1-AV-incongruent	2	1.98 (0.85)	0.12 (0.14)	88 %
2	2-AV-count flashes	2	2.01 (0.88)	0.13 (0.17)	88 %
2	2-AV-count events	2	2.00 (0.83)	0.04 (0.06)	96 %
2	2-AV-count beeps	2	2.03 (0.87)	0.07 (0.12)	92 %
3	3-V	5.5	3.77 (1.75)	0.54 (0.54)	27 %
3	3-A	5.5	4.00 (1.94)	0.55 (0.93)	32 %
3	3-AV	5.5	4.23 (1.98)	0.56 (1.23)	33 %

The third columns shows the mean response of a perfect observer. The fourth column shows the mean response and SD (between brackets) of all participants. The fifth column shows the average of the variance in responses per participant (with SD between brackets). The last columns shows the percentage of correct responses. Note that the data for the multisensory conditions do not include the incongruent data.

First hypothesis

The first hypothesis stated that the presence or absence of incongruent stimuli caused or contributed to the discrepancy. The following findings would support the first hypothesis:

- the mean of the 1-AV-congruent condition is larger than the mean of the largest unisensory condition (i.e., 1-A), and
- the mean of the 1-AV-incongruent condition is not larger than the mean of the largest unisensory condition, and
- the variance of the 1-AV-congruent condition is larger than the variance of the smallest unisensory condition (i.e., 1-A), and
- the variance of the 1-AV-incongruent condition is smaller than the variance of the smallest unisensory condition.

Statistical analyses revealed an effect of Condition on the variance ($F_{3,27} = 5.92$, $p < .01$), but not on the mean ($F_{3,27} = 0.65$, $p = .58$). The post hoc showed that only the variances between the 1-A and the 1-V did differ significantly though (*p < .01)*. So, we found no differences between the mean or the variance of the 1-AV-congruent and the 1-A condition and thus no support for the first hypothesis.

Second hypothesis

The second hypothesis stated that the instruction to the observer caused or contributed to the discrepancy. The following findings would support the second hypothesis:

- the mean of the 2-AV-count events condition is larger than the mean of the largest unisensory condition (i.e., 1-A), and
- the mean of the 2-AV-count flashes condition is not larger than the mean of the largest unisensory condition, and

- the variance of the 2-AV-count events condition is larger than the variance of the smallest unisensory condition (i.e., 1-A), and
- the variance of the 2-AV-count flashes condition is not larger than the variance of the largest unisensory condition.

Statistical analyses of the 1-A, 2-AV-count events, 2-AV-count flashes and 2-AV-count beeps revealed an effect on Condition on the variance ($F_{4,18} = 3.89$, $p < .001$), but not on the mean ($F_{4,18} = 0.64$, $p = .64$). The post hoc showed that the variance of the 2-AV-count events condition did not differ from the 1-A condition ($p = 1.00$), but was significantly *smaller* than the 1-V condition ($p < .01$). The variance of the 2-AV count flashes condition did not differ from both the 1-A ($p = 0.06$) and 1-V ($p = 1.00$) conditions, but was significantly *larger* than the variance of the 2-AV count events condition ($p < .05$). These findings do not support the second hypothesis.

Third hypothesis

The third hypothesis stated that the range of the number of pulses caused or contributed to the discrepancy. If so, the results of Experiment 4.3 should match the effects previously reported in TNJ, while the results of Experiment 4.1 should match the effects previously reported in the FI paradigm. In other words, the following findings would support the third hypothesis:

- the mean of the 3-AV condition is larger than the mean of the largest unisensory condition (i.e., 3-A), and
- the mean of the 1-AV-congruent is not larger than the mean of the largest unisensory condition (i.e., 1-A), and
- the variance of the 3-AV condition is larger than the variance of the smallest unisensory condition (i.e., 3-V), and
- the variance of the 1-AV-congruent condition is smaller than the variance of the smallest unisensory condition (i.e., 1-A).

Statistical analyses showed that a significant effect of Condition on the mean ($F_{2,18} = 4.26$, $p < .05$). Specifically, the 3-AV condition differed from both the 3-A and the 3-V conditions (both $p < .001$). There was no significant effect of Condition on the variance ($F_{2,18} = 0.01$, $p = .98$). Thus, the differences in the mean support the third hypothesis.

Figure 4.2 displays the mean reported number of pulses as a function of the presented number of pulses. This figure suggests that observers tend to report a higher number of perceived pulses in the 3-AV condition for the whole range of pulses presented, including the range of 1-3. To further investigate these effects, we separately analyzed the results for 1 to 3 pulses in Experiment 4.3. For this range, the basic statistics were (mean with SD and variance with SD): 1.92 (0.89) and 0.29 (0.40) for 3-V, 1.96 (0.85) and 0.20 (0.40) for 3-A, and 2.13 (0.95) and 0.21 (0.22) for 3-AV. Statistical analyses on the range 1-3 showed an effect of Condition on the mean ($F_{2,18} = 4.26$, $p < .05$), but not on the variance ($F_{2,18} = 0.32$, $p = .72$). Specifically, the mean of the 3-AV condition differed from both the 3-A and 3-V conditions (both $p < .001$). So, also for the judgment of 1 to 3 pulses, multisensory presentation affected the mean in Experiment 4.3.

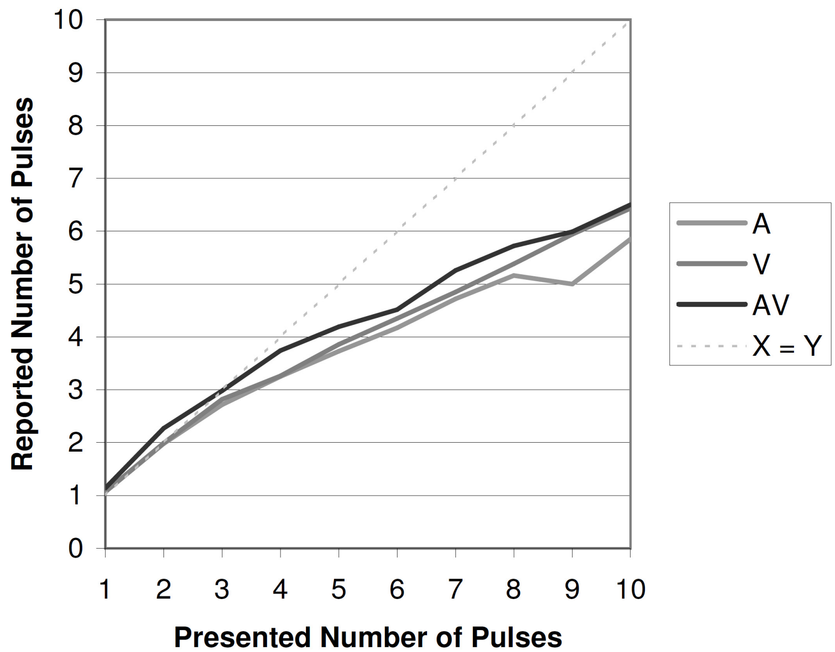


Figure 4.2: Mean numerosity functions in Experiment 4.3. The mean reported number of pulses is displayed as a function of the presented number of pulses for each of the three conditions. The three solid lines show the mean responses for the auditory (gray), visual (dark gray) and audiovisual (black) conditions. The dashed line represents the mean response of a perfect observer (light gray).

4.4 Discussion

For the counting of pulses in rapidly presented series, the effects of multisensory presentation on the mean (and to a lesser extend on the variance) of numerosity estimates depend on the experimental paradigm (viz. the FI or the TNJ paradigm). The objective of this study was to investigate the three differences between both paradigms as a cause for this discrepancy: the *presence or absence of incongruent stimuli*, the *instruction to the observer*, and the *range of the number of pulses*. The general results of the three experiments are in line with the general results of both FI studies and TNJ studies. Just as in previous FI studies, participants who were presented with congruent series of a small number of pulses could count them accurately. Consistent with the reports in TNJ, we find that participants tend to underestimate the number of pulses presented when their number is large, but not when their number is small. Also, we find no multisensory effects on the variance of the numerosity estimates. However, in both FI and TNJ an effect on the variance is only rarely reported, indicating that if it exists, it is probably small and difficult to reproduce.

Differences in pulse range causes the discrepancy in multisensory effects in FI and TNJ

In Experiment 4.1, we investigated the hypothesis that the presence (as in FI) or absence (as in TNJ) of incongruent stimuli caused or contributed to the discrepancy. We found no differences between the numerosity estimates of the multisensory stimuli in the congruent condition and the unisensory stimuli. Thus, we find no support for this hypothesis. Likely, the numerosity estimates of the congruent stimuli were not affected by the presentation of incongruent stimuli, because the estimates of the congruent stimuli were relatively accurate whereas the estimates of incongruent stimuli are usually more inaccurate. Accurate estimates are usually not affected by inaccurate estimates (Andersen *et al.,* 2004; Bresciani & Ernst, 2007; Hotting & Roder, 2004). The absence of a favorable multisensory effect on the variance may be explained by the facts that 1) observers' estimates had a larger variance when counting flashes than counting beeps and that 2) observers were instructed to count the number of flashes in the multisensory stimuli.

In Experiment 4.2, we investigated whether the instruction to the observer caused or contributed to the discrepancy. In comparison with the unisensory estimates, we found a favorable multisensory effect on the variance when participants are instructed to count events (as in TNJ) and not when they are instructed to count flashes (as in FI), but no effect on the mean estimate. The effects in variance were opposite to what we expected. In previous studies, favorable multisensory effects on the variance were only reported for the FI paradigm, while favorable effects on the mean are a typical result for the TNJ paradigm. This means that the different instructions to the observer cannot explain the discrepancy between both paradigms. The difference in variance between when observers were counting events and when they were counting beeps may be explained by that the instruction to count events actually encourages participants to integrate information (which may reduce the variance).

In Experiment 4.3, we investigated whether the range of the number of pulses caused or contributed to the discrepancy. When a large range of pulses was presented (i.e., in Experiment 4.3), we found an increase in the mean numerosity estimates of the multisensory series in comparison with the unisensory series. However, this difference was absent when a small range of pulses was presented (i.e. in Experiment 4.1). These results thus support the third hypothesis.

How can different pulse ranges affect multisensory effects?

Now that we can conclude that the difference in multisensory effects in the FI and TNJ is caused by the different range of pulses presented in both paradigms, the question is which mechanism can cause this effect. Based upon the results in the previous chapter, we hypothesized that the multisensory effect is proportional to the number of pulses presented, and thus larger for larger number of pulses. This means that the multisensory effects for the large and small range would differ because the large range contains more stimuli with larger numbers of pulses than the small range. However, this hypothesis is invalidated by the fact that the effects remain present for small pulse numbers in the large range (i.e. in Experiment 4.3; see Figure 4.2), while they are absent for small pulse numbers in the small range (i.e. in Experiment 4.1). This means that the mechanism we proposed to explain how multisensory presentation affects the mean estimate in the previous chapter may not be correct.

An alternative explanation for the data in this chapter may lie in the strategies employed to determine the number of pulses in large and small ranges. Given the rapid presentation speed of pulses, participants may be able to count the number of pulses when their number is small, but are forced to estimate when their number is large, for instance based on the amount of energy in a series (see Kaufman, Lord, Reese & Volkmann, 1949). Since the presentation rate of the pulses is constant, a valid strategy would be to estimate the duration of a series. If participants are not able to switch quickly between strategies, they may also use duration estimation for small numbers of pulses although these could also be counted. This leads to the use of different strategies for 1-3 pulses that depends on the different ranges in a block (viz. counting if the total range is 1-3 and duration estimation if the range is 1-10). The question then is: why does multisensory presentation have different effects on counting and duration estimation?

Different multisensory effects on counting and duration estimation

(Interval) duration can be influenced by non-temporal characteristics of the marker and its modality, nature (i.e. filled vs. non-filled intervals and adjacent stimuli), and its energy (Allen, 1979; Grondin, 2003; Van Erp & Spapé, 2008; Van Erp & Werkhoven, 2004). According to the theories of time perception, an accumulator/pacemaker produces pulses at a constant rate which can be counted or stored for comparison (Treisman, 1963; Gibbon, Church & Speck, 1984). It is thought that the above characteristics can affect the speed at which the accumulator/pacemaker releases pulses, with higher speeds leading to the perception of longer durations. For example, Wearden and colleagues (1998) argue that auditory stimuli have a longer perceived duration than visual stimuli because auditory stimuli induce a higher accumulator/pacemaker speed than visual stimuli. This is in accordance with the findings in Experiment 4.3, where participants estimated the number of pulses in auditory series higher than in visual series.

Multisensory audio-visual stimuli may similarly increase the accumulator/pacemaker speed in comparison to unisensory stimuli. The speed of the pacemaker could be higher because it is stimulated by both audition and vision or because it is stimulated by more pulses. This hypothesis is compatible with several studies exploring audiovisual interactions in the perception of event duration, which show that either sound can extend the perceived duration of a visual event (Burr, Banks & Morrone*,* 2009; Chen & Yeh, 2009; Klink, Montijn & Van Wezel, 2011; Walker & Scott, 1981) or that vision can extend the perceived duration of an auditory event (Van Wassenhoven, Buonamono, Shimojo & Shams, 2008). Recently, Chen and Yeh (2009) explained that these effects are due to an increase in the pulse production rate of the pacemaker. More recently, Klink and colleagues (2011) showed that they could also be caused by alteration of the on- and offset of the pacemaker itself. However, this second mechanism may not have played a role in our experiment, because it required the presentation of asynchronous auditory and visual events.

Conclusion

We conclude that the range of pulses presented – and not the presence or absence of incongruent stimuli, or the instruction to the observer – causes the different multisensory effects found for the FI and TNJ paradigms. For the small pulse range as in FI, there is no effect of congruent multisensory stimuli because observers use a mechanism that is less sensitive to multisensory presentation, namely accurate counting. For the large pulse range as in TNJ, accurate counting is not possible and observers switch to a strategy based on duration estimation. For the large pulse range, there is an effect of congruent multisensory stimuli because observers use a mechanism that is sensitive to multisensory presentation, namely duration estimation.

Chapter 5. Visuotactile encoding and retrieval of object identity and object location[4]

Abstract

Researchers have reported that audiovisual object presentation improves memory encoding of object identity in comparison to either auditory or visual object presentation. However, multisensory memory effects on retrieval, on object location, and of other multisensorial combinations are yet unknown. We investigated the effects of visuotactile presentation on the encoding *and* retrieval of object identity memory *and* object location memory. Participants completed a memory test consisting of an encoding and retrieval phase. In the encoding phase (c) they explored four game-like cards presented on a computer screen in a two by two arrangement. Participants could touch each card to experience its content: a Morse code presented on the screen (V) and/or via a tactile vibrator attached to the participant's index finger (T). In the retrieval phase (r), they had to indicate for each of eight cards if (recognition) and where (relocation) it had been presented earlier. Compared with the visual base line (cV-rV), we found that both 'multisensory encoding' (cVT-rV) and 'multisensory retrieval' (cV-rVT) significantly improved both recognition and relocation performance (all at least $p < .05$). Compared with the tactile base line (cT-rT), we found no multisensory encoding or retrieval effects ($p = .79$ and $p = .85$). We conclude that visuotactile presentation can improve memory encoding *and* retrieval of object identity *and* location. However, it is not yet clear whether these benefits are due to multisensory interactions or simply due to the multiple encoding and retrieval operations initiated by multiple sensory modalities.

---

[4] Parts of this chapter have been submitted as:

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. (*under revision*). Visuotactile memory of object identity and object location.

5.1 Introduction

The effect of multisensory stimulus presentation on human performance has been studied extensively. Compared with unisensory presentation, multisensory presentation can positively affect *perception* (Hershenson, 1962; Ernst & Bülthoff, 2004, Philippi, van Erp & Werkhoven, 2008; Stein & Meredith, 1993), *attention* (Driver & Spence, 1998; Spence & Driver, 2004; Werkhoven, van Erp & Philippi, 2009) and *memory* (Delogu, Raffone & Olivetti Belardinelli, 2009; Shams & Seitz, 2008; Thompson & Paivio, 1994). However, the majority of the multisensory research efforts are focused on multisensory perception while only a limited number of studies investigated multisensory effects on higher cognitive processes like memory. Actually, to our knowledge, the understanding of multisensory memory is currently limited to the effects of audiovisual presentations on encoding (but not retrieval) in object recognition (but not localization) tasks.

Audiovisual encoding of object identity

Thompson and Paivio (1994) were one of the first to report that multisensory stimuli were better memorized than unisensory stimuli. They serially presented twenty sounds, twenty pictures, or twenty sound-picture pairs and asked participants to recall as many objects as they could. The results showed that the number of objects recalled in the multisensory condition was approximately 50% higher than in each unisensory condition. Thompson and Paivio explained these results with the dual coding theory (Paivio, 1971; 1986). According to this theory, a stimulus that is presented to multiple modalities is encoded once in each modality and may be recalled via either modality, which leads to superior recall.

Recently, Murray and colleagues (Lehmann & Murray, 2005; Murray *et al.,* 2004; Murray, Foxe & Wylie, 2005) reported multisensory memory benefits in a continuous recognition task. Pictures were recognized about 5% more often as presented earlier when they had been presented earlier with a matching sound than when they had been presented earlier without or with a non-matching sound. Likewise, Von Kriegstein and Giraud (2006) reported that auditory speaker recognition improved by about 10% when participants were earlier exposed to voice-face pairs instead of voices only. These findings suggest that multisensory object presentation improves the encoding in memory. Shams and Sietz (2008) argue that these multisensory memory effects are the result of multisensory interactions. When multiple sensory modalities process information about a single object or event, the information processing in each of those sensory modality may interact with (i.e., alter) the information processed in other sensory modalities. These interactions may change how we perceive that object or event and/or affect how (well) we encode it (see also Murray *et al.,* 2004). It should be noted though, that this explanation does not rule out the explanation provided by Thompson and Paivio. That is, the perception of a multisensory stimulus could initiate multiple encoding processes, which are each enhanced via sensory interactions.

Expanding to retrieval

Adequate memory performance relies not only on encoding information in memory but also on retrieving information from memory. So far, multisensory memory studies either assessed multisensory effects on encoding or have not distinguished between both processes. Therefore, the first goal of this study is to investigate multisensory effects on the retrieval process in addition to those on the encoding process. Studies investigating the relation between encoding and retrieval (Slotnick, 2004; Kent & Lambers, 2008) indicate that similar cortical circuits are activated during encoding and retrieval. This may mean that benefits of multisensory encoding may be similar to benefits of multisensory retrieval. Consequently, we hypothesize that multisensory presentation can improve retrieval. A multisensory stimulus is processed by multiple sensory modalities. Each of these modalities may independently attempt to retrieve information from memory, which may increase memory performance in comparison to a unisensory retrieval cue (probability summation). This idea is similar to Thompson and Paivio's explanation of multisensory encoding, in the sense that it considers each sense to initiate an independent memory process. In addition, sensory interactions, which may occur during the processing of a multisensory cue, could increase retrieval effectiveness by reducing the signal to noise ratio of the retrieval process. This idea is similar to the explanation that multisensory interactions alter perceptual/memory processing (Shams and Seitz (2008). Both mechanisms are schematized in Figure 5.1.

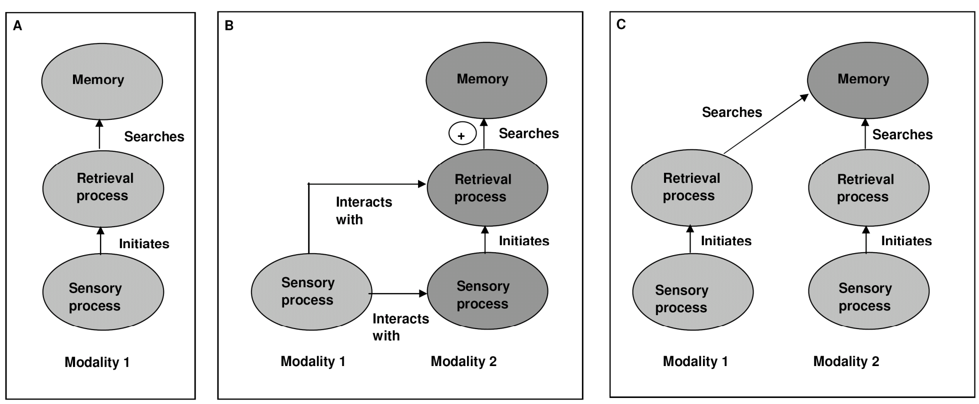


Figure 5.1. Uni- and multisensory retrieval mechanisms. Panel A details the unisensory retrieval of information. *Sensory processing* of a unisensory stimulus initiates a *retrieval process*. *Memory* is viewed as a database containing information. The *retrieval* process compares information from the unisensory stimulus with that in memory. Successful retrieval allows an observer to recognize an object and/or its location. When multiple sensory modalities process information two possible mechanisms may boost retrieval effectiveness. First, in panel B, the information perceived in the second modality may improve sensory processing in the other modality and/or improve the retrieval of information initiated by the second modality (for instance, the presence of a sound in the second modality may limit searching information in memory to objects who produce sound). Second, in panel C, each sensory modality that processes information may initiate its own *retrieval process.* The mechanisms in panels B and C are not mutually exclusive (see panel D of Figure 1.5 for an illustration that combines both mechanisms).

Expanding to touch and object location memory

Currently known effects of multisensory memory are also limited to the effects of audiovisual presentation on object identity. Studies in multisensory perception have shown that beneficial multisensory effects extent to combinations with touch (Bresciani, Dammeier & Ernst, 2006; Bresciani *et al*., 2005, Hötting & Röder, 2004; Violentjev, Shimojo & Shams, 2005) and to spatial dimensions (Alais & Burr, 2004; Ernst & Banks, 2002). This raises the questions if multisensory memory effects can similarly be generalized to modality combinations with touch and if multisensory memory benefits exist for object location. Our second and third goal are to investigate visuotactile effects on memory performance and to investigate multisensory memory for object identity as well as object location. Based upon the findings in multisensory perception and attention research, we expect that visuotactile presentation can improve memory performance for both object identity and object location tasks. Multisensory memory effects for object location, however, may not have the same size as those for object identity. Recently, Lehnert and colleagues (Lehnert & Zimmer, 2006; 2008) investigated identity and location memory for audition and vision. Their results indicate that auditory and visual spatial memory is commonly coded, while auditory and visual object memory is modality-specific. This implies that different multisensory benefits may occur for the encoding and retrieval of object identity and object location.

Memory card game with Morse codes

To investigate these hypotheses we employed an electronic memory card game in which the memory cards contained simple Morse code. We choose to present Morse code because of two reasons. First, Morse code can be presented easily in both the visual and tactile modalities. Second, the Morse code presented to each modality is *redundant*. This means that the same information (viz., a temporal sequence) is presented to each sense. In contrast, semantic information presented to each modality is usually *complementary*. Complementary information presented to each sense refers to the same object, but the information actually presented is not the same (see Ernst & Bülthoff [2004] for a more elaborate discussion on complementary and redundant sensory information). Presenting redundant stimuli has the advantage that any multisensory effect must be due the presentation of *information in multiple modalities* whereas multisensory effects for complementary stimuli may also be due to the presentation of *more information*.

5.2 Method

Participants

Eighteen (14 male, 4 female) students with ages between 21 and 35 years (mean = 24.3; SD = 3.6) from Utrecht University participated in the experiment. All had normal or corrected-to-normal vision and none reported any experience with Morse code or a history of touch disorder. They gave their informed consent prior to participation.

Setup

Participants were seated in front of a 17 inch CRT touch screen (Elo, 1024x768 pixels, 85 Hz frequency). The viewing distance was approximately 55 cm. A custom build tactile vibrator (TNO JHJ-3, see Van Erp et al., 2007) was attached to their right index finger. To mask the sound produced by the vibrator, participants wore noise protection headphones (Bilsom 747).

Stimuli

The touch screen displayed the identical backsides of the memory cards. When a participant touched a card with his or her right index finger, 500 ms later the card's object (i.e., the Morse code) was presented visually on the monitor and/or tactilely with a tactile vibrator on the same finger. Participants were instructed to keep touching the card while its object was presented; early release terminated object presentation.

The Morse codes were presented as on-off signals with durations of 200 ms ('dots') and 600 ms ('dashes') separated by intervals of 200 ms. This is equivalent to the Paris standard of 6 words per minutes. All Morse codes presented in this experiment consisted of one to three dots and/or dashes (14 codes in total) and had total durations between 200 ms (a single dot) and 2200 ms (three dashes). The visual dots and dashes were presented by a flashing filled white disc with a diameter of 2 degrees of visual angle against a dark background at the location of the touched memory card. Tactile dots and dashes were presented as 160 Hz bursts of vibration to the right index finger for the appropriate duration. When visual and tactile dots and dashes were presented simultaneously, their on- and offset was synchronous.

Control experiment

Simultaneous presentation of temporal patterns in multiple sensory modalities affects perception. As we described in Chapters 3 and 4, observers may report less underestimation when they estimate a series of multisensory pulses than when they estimate a series of unisensory pulses (Philippi, Van Erp & Werkhoven, 2008). A (beneficial) multisensory effect on perception can indirectly affect memory performance. Therefore, we ensured that the Morse codes were perfectly perceivable in each sensory modality. Prior to the experiment, we therefore tested a different group of 17 participants (9 male, 8 female) with ages between 20 and 35 (SD = 2.8). They were twice presented with 42 memory cards containing the objects presented in the main experiment and were instructed to touch each card once and write down its code in dots and dashes on a sheet of paper. Half of the participants first completed 42 memory cards with visual objects and completed the other 42 memory cards with tactile objects, while the other half started with tactile objects and finished with visual objects. On average, participants reported 41.2 out of 42 visual objects (SD = 1.33) correctly, and 41.8 out of 42 tactile objects (SD = 0.44). A Pearson's Chi-square test showed that both the visual ($\chi^2 = .96$; $p = 1.00$) and the tactile score ($\chi^2 = .10$; $p = 1.00$) did not differ from a perfect score.

Table 5.1: The presentation modalities for each phase in the seven experimental conditions

Phase		Encoding (c)		
	Presentation Modalities	Vision (V)	Touch (T)	Vision and Touch (VT)
Retrieval (r)	Vision (V)	cV-rV	-	cVT-rV
	Touch (T)	-	cT-rT	cVT-rT
	Vision and Touch (VT)	cV-rVT	cT-rVT	cVT-rVT

For each condition (cells) it is listed in what modality or modalities cards were presented in the encoding (columns) and retrieval (rows) phase of each game. Cross-sensory conditions were not tested because they were not relevant to the goals set in this paper.

Conditions

Each memory game consisted of an encoding (c) and a retrieval (r) phase. The experiment had seven experimental conditions, which differed in which modalities the card's objects were presented in each phase (V, T, VT; see Table 5.1). These experimental conditions were chosen to investigate the three goals set in this study.

The first goal was to investigate the multisensory effects on retrieval. To achieve this goal we compared the visual baseline (cV-rV) and the tactile baseline (cT-rT) with their related multisensory retrieval conditions (respectively, cV-rVT and cT-rVT).

The second goal was to investigate multisensory memory effects for the combination of vision and touch. To achieve this goal we compared the visual and tactile baseline with their respective multisensory encoding conditions (cVT-rV) and (cVT-rT). These comparisons are similar to those conducted on audiovisual memory by Murray and colleagues (Lehmann & Murray, 2005; Murray *et al.* 2004) and by Von Kriegstein and Giraud (2006). Secondly, we also compared the visual and tactile baseline with the multisensory encoding and retrieval condition (cVT-rVT). These comparisons mirror those conducted by Thompson and Paivio (1994).

Finally, the third goal was to investigate multisensory memory effects in object location memory. This was done by conducting the comparisons described above on the object location data instead of the object identity data.

Design and procedure

Each participant completed eight blocks of seven memory games. In each block, they played one game in each experimental condition. The order of experimental conditions was balanced within participants across blocks and across participants.

Each memory game started with an encoding phase. In each encoding phase, four memory cards were presented as displayed in the left panel of Figure 5.2. Participants could explore these cards for 20 seconds and were instructed to remember the content and location of each card. In addition, participants had to verbally repeat the letter string ‘a,b,c’ during the encoding phase. This articulatory suppression task was used to prevent verbal recoding of the object identities and locations (see Thompson & Paivio, 1994).

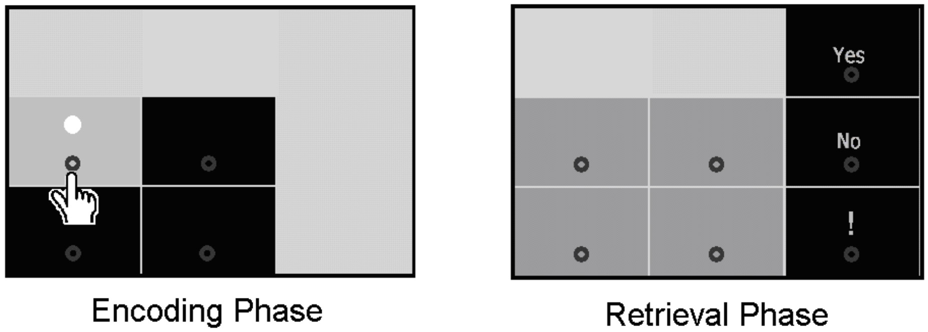


Figure 5.2 – Screens of the encoding and the retrieval phase. These show the arrangement of cards and buttons during the encoding and retrieval phases. In the encoding phase (left panel) four memory cards were presented and a participant could touch each one to experience its Morse code. In the retrieval phase (right panel), participants could experience a Morse code by pressing the exclamation card button. Thereafter they could indicate if and where this code was presented in the encoding phase by pressing either the ‘Yes’ or the ‘No’ button and an appropriate location card.

At the end of the encoding phase all cards disappeared. Two seconds later, the retrieval phase started. In the retrieval phase, four grayed out location cards, an exclamation mark button, a ‘Yes’ button, and a ‘No’ button were displayed on the touch screen, as shown in the right panel of Figure 5.2. When the participants touched the exclamation mark button, it was replaced by a test card which presented its object in the appropriate modality. Once the object presentation was finished, the test card was replaced by a question mark and participants had to indicate if and where the test card’s object was presented in the encoding phase by first touching either the ‘Yes’ or ‘No’ button and then by touching a location card. Note that even when a participant indicated that an object had not been present in the encoding phase he or she still had to give a location: participants were instructed to give the location of the object with the most resemblance to the object they just observed. This procedure was adopted so that the object identity data and the object location data were independent. Participants were tested with eight test cards in each retrieval phase, which were presented in random order. Of these cards, four had been presented in the encoding phase (targets) and four had not (non-targets). The four non-targets resembled the targets in the number of dots and dashes they contained and in whether they contained both dots and dashes or either dots or dashes. Participants were encouraged to not think for too long about the identity and location of a single object because that may impede their ability to remember further objects.

The objects presented in the encoding and retrieval phase differed per game. Pilot studies indicated that the fewer items a Morse code contained, the easier it was to remember. Morse codes containing either only dots or only dashes were also easier to remember than Morse codes containing both dots and dashes. The following measures were taken to equalize the difficulty of different games:

- Each game contained zero, or one Morse code consisting of one dot or dash.

- Each game contained zero, one, or two Morse codes consisting of two dots or dashes.
- Each game contained two or three Morse codes consisting of three dots or dashes.
- Each game contained one or two Morse codes consisting of only dots or only dashes.

The individual Morse codes presented in each game were balanced within participants and experimental conditions.

5.3 Results

In the retrieval phase, participants took 4.5 s (SD = 1.4) on average to resolve each test card. Specifically, once participants touched the exclamation mark button, 1.8 s (SD = 0.6) elapsed to display the object. They then took 2.2 s (SD = 1.3) to touch the Yes or No button and 0.5 s (SD = 0.4) to touch a location button. The average time to complete a retrieval phase was 40.8 s (SD = 7.8).

The object identity data showed that participants correctly identified 2.5 targets (SD = 1.1) and 2.7 non-targets (SD = 1.0) per game. To control for response biases in the object identity data we applied Signal Detection Theory (see Green & Swets, 1966) and used the sensitivity index *d′* (d-prime) as a measure of the separation between the means of the signal and the noise distributions, in units of the standard deviation of the noise distribution. The index *d′* can be computed from measurements of the hit rate (correctly identified) and false-alarm rate (incorrectly identified) and is calculated as *d′* = Z(hit rate) - Z(false alarm rate), where function Z is the inverse of the cumulative Gaussian distribution. In this experiment the *d'* could attain values between -2.56 and 2.56. In the object identity task, participants scored above chance level (i.e., *d'* > 0) in 64.9% of the games played.

For the object location data, we computed the *relocation score* as the fraction of targets correctly relocated for each game, ranging from 0 to 1. In the object location task, the average relocation score was 0.53 (SD = 0.31). Participants scored above chance level in 66.2% of the games played (i.e., 2 or more correctly relocated items).

Multisensory retrieval effects

Multisensory retrieval effects were assessed by comparing the object identity data in the cV-rV with the cV-rVT condition and the cT-rT with the cT-rVT condition. The *d'* scores for these conditions, as well as the *d*'s for the multisensory encoding conditions (which are discussed below), are displayed in Figure 5.3. Two separate repeated measures ANOVAs (one assessing multisensory effects related to the visual baseline and the other related to the tactile baseline) were performed, each with Experimental Condition (4 levels) as independent variable. We found a significant main effect for the multisensory effects related to the visual baseline ($F_{3,51} = 4.17$; $p < .05$) but not related to the tactile baseline ($F_{3,51} = 0.33$; $p = .79$). A Fisher post-hoc analyses was conducted, whose results are displayed in Figure 5.3. They show a multisensory retrieval effect, but only when touch was added to vision during retrieval.

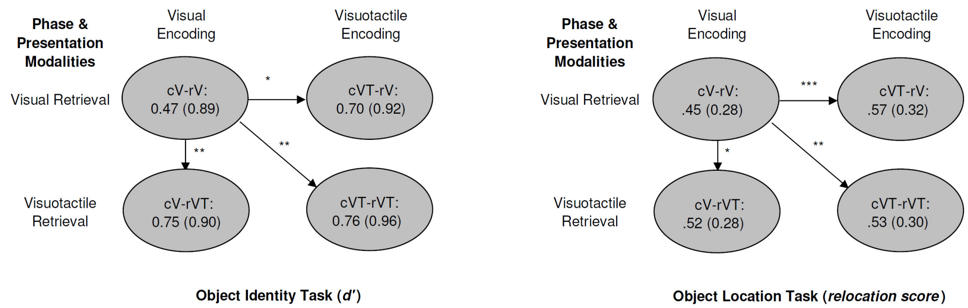


Figure 5.3 – Memory performance in the object identity task. The left section displays the multisensory encoding and retrieval effects related to the visual baseline. The right section displays the multisensory encoding and retrieval effects related to the tactile baseline. Mean *d's* are displayed without brackets and standard deviations are displayed within brackets. Significant differences are denoted by one ($p < .05$), two ($p < .01$), or three asterisks ($p < .001$).

Multisensory memory (encoding) effects for vision and touch

Next, we investigated whether multisensory (encoding) effects reported earlier for audition and vision extend to the combination of vision and touch. For this, we compared the object identity data of 1) the visual baseline (cV-rV) with the relevant multisensory encoding (cVT-rV) and multisensory encoding and retrieval (cVT-rVT) conditions, and 2) the tactile baseline (cT-rT) with the relevant multisensory conditions (cVT-rT; cVT-rVT). These comparisons were conducted as part of the analyses of the previous paragraph; the results are shown in Figure 5.3. Multisensory encoding improves object identity memory, but only when touch was added to vision.

Multisensory memory effects on object location

Figure 5.4 displays the *relocation scores* for all experimental conditions. Just as with the analyses of the object identity data two separate repeated measures ANOVAs were conducted to investigate the multisensory memory effects on object location. A significant main effect was found for multisensory effects related to the visual baseline ($F_{3,51} = 3.53$; $p < .05$), but not for multisensory effects related to the tactile baseline ($F_{3,51} = 0.26$; $p = .85$). The results of a subsequent Fisher LSD post-hoc test are displayed in Figure 5.3. They show that multisensory encoding and retrieval effects also exist for object location memory.

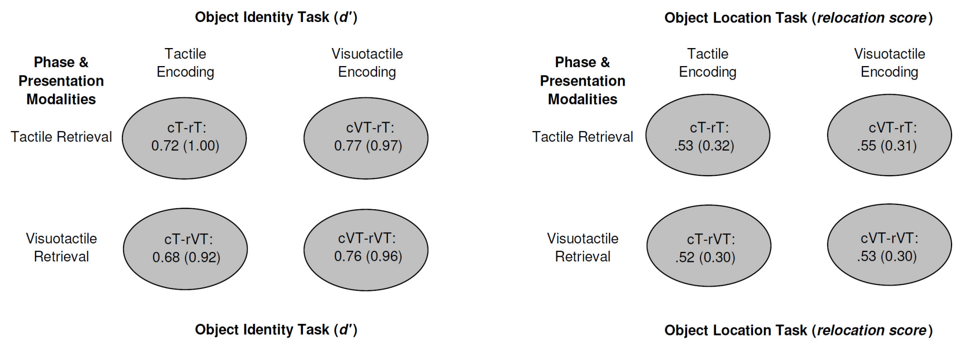


Figure 5.4 – Memory performance in the object location task. The left section displays the multisensory encoding and retrieval effects related to the visual baseline. The right section displays the multisensory encoding and retrieval effects related to the tactile baseline. Mean *d'* are displayed without brackets and standard deviation are displayed within brackets. Significant differences are denoted by one ($p < .05$), two ($p < .01$), or three asterisks ($p < .001$).

5.4 Discussion

We investigated the effects of visuotactile presentation on memory encoding and memory retrieval of object identity and object location. We presented visual, tactile and visuotactile objects and found effects of multisensory presentation on encoding as well as retrieval in object identity and object location memory. These effects confirm our hypotheses. The multisensory presentation effects we report are, however, limited to the benefits of adding touch to vision; we find no effects for adding vision to touch at all. We will further discuss this asymmetry and its consequences after we reflect on the three hypotheses.

Benefits of multisensory retrieval

A comparison between the visual baseline (cV-rV) and the relevant multisensory retrieval condition (cV-rVT) reveals a significant benefit for presenting visuo-tactile presentation during retrieval. This indicates that a multisensory cue is more effective in retrieving information than a unisensory cue. The comparison also reveals that more information is encoded during visual encoding (cV) than is retrieved by visual retrieval cues (rV). In other words, visual retrieval did not access all available information (see also Tulving & Pearlstone, 1966). Furthermore, the effect of multisensory retrieval is similar the effect of multisensory encoding (cV-rV with cVT-rV). This finding is consistent with the idea that similar neural circuits are activated during encoding and retrieval (Slotnick, 2004).

As proposed in the introduction, this multisensory retrieval effect may be achieved via two not mutually exclusive mechanisms (see Figure 5.1). First, the unisensory components of a multisensory retrieval cue could each initiate an attempt to retrieve information independently, followed by probability summation. Second, as a result of multisensory integration of redundant information, the unisensory components of a multisensory retrieval cue could interact and improve the signal to noise ratio of each retrieval attempt. As shown in Figure 5.1c, this second mechanism may hold even if multisensory interactions do not affect perception.

Based upon our data we cannot distinguish between these two mechanisms. A possible approach to disentangle these mechanisms is to study multisensory memory effects for objects whose unisensory components are presented from different locations at different times. When the unisensory components of a single object are not presented simultaneously, each sensory modality may still initiate its own retrieval (or encoding) process, but no multisensory interactions should occur.

Benefits of multisensory memory (encoding) for touch and vision

As mentioned in the previous paragraph, we find a significant benefit of multisensory encoding (cVT-rV) in comparison with the visual baseline (cV-rV). For object identity memory, adding touch to vision during encoding increases sensitivity by about 70%, which roughly corresponds to an increase in the number of hits and correct rejections of 10%. This effect is in line with the performance benefits reported for audiovisual encoding (cAV-rV) in comparison with a visual baseline (cV-rV) as reported by Murray and colleagues (Lehmann & Murray, 2005; Murray *et al*. 2004), and by Von Kriegstein and Giraud (2008). Murray and colleagues conducted continuous picture recognition tasks and found that pictures were recognized more often (by about 5%) when they had earlier been presented with a matching sound. Von Kriegstein and Giraud reported an 10% improvement in auditory speaker recognition when participants were earlier briefly exposed to voice-face pairs instead of voices only. Possibly, the benefits of adding touch to vision and the benefits of adding sound to vision are achieved via similar mechanisms.

To our surprise, we find no additional effect of combining multisensory encoding and multisensory retrieval. A possible explanation is that the retrieval of encoded information is a non-linear process, in which an improvement in encoding or retrieval may not always lead to a (measurable) improvement in memory performance. For example, if multisensory presentation affects encoding such that subsequent retrieval operations always succeed, then no benefit should be expected for multisensory retrieval.

Benefits of multisensory object location memory

There is large resemblance in the object identity and object location memory data. The resemblance can be explained by a hypothesis coined by Hasher and Zacks (1979). They hypothesized that object locations are encoded automatically into memory due to their ecological significance. Under this hypothesis, any improvement in object identity memory would also result in an increase in object location memory. The Hasher and Zacks hypothesis has, however, not been supported consistently (see Postma *et al*., 2008).

Postma and colleagues (Postma *et al*., 2004; 2008) consider object identity memory and object location memory to be independent. We propose that the benefit of multisensory object identity presentation on object identity memory is equal to the benefit of multisensory location presentation on object location memory. Future studies may want to investigate the effects of multisensory presentation on object identity memory and object location when only the identity or only the location is presented to multiple sensory modalities.

Assymmetries in the effects of adding touch to vision and vice versa

Overall, the results indicate that it is easier to recognize and relocate Morse codes when they are presented tactilely than when they are presented visually. Because we made sure that the objects presented were equally well perceivable in both modalities this difference must be due to a difference in encoding and/or retrieval through vision and touch and not due to a difference in perception. This finding could be considered an extension of earlier studies that showed that touch was more accurate than vision in perceiving temporal patterns (Lechelt, 1975; Philippi, Van Erp & Werkhoven, 2008; Van Erp & Werkhoven, 2004).

A reverse asymmetry was reported for visual and tactile recognition memory following the exploration of physical objects (Newell, Woods, Tjan & Bulthoff; Woods & Newell, 2004). That is, a longer presentation duration was required for tactile exploration than for visual exploration to reach the same level of recognition performance. However, this was likely caused by the fact that tactilely only (a part of) one object could be explored at a time while visually multiple objects could be explored simultaneously. Thus, this difference may not have much to do with differences in encoding and retrieval of tactile and visual memory.

No asymmetry between auditory and visual memory was found by Thompson and Paivio (1994). They presented sounds, pictures, or sound-picture pairs and found that the number of recalled sounds and pictures was equal. In addition, they also found that the number of recalled sound-picture pairs was higher than the number of sounds and the number of pictures, which also contrasts our (asymmetric) multisensory memory effects. Future studies may want to investigate these differences by systematically comparing the differences between the study by Thompson and Paivio and ours. Possibly, difference in multisensory benefits occurs because Thompson and Paivio's sound-picture pairs carried semantic information whereas Morse code did not (at least not for our participants). Semantics are known to play an important role in multisensory object recognition (Laurenti, Kraft, Madljian, Burdette & Wallace, 2004; Yuvall-Greenberg & Deouell, 2007; 2009).

The mechanism underlying multisensory memory benefits

The absence of any effect of touch on vision raises questions about the nature of the multisensory presentation effects reported in this study. Based upon the data gathered in this experiment it is tempting to assume that memory performance increases when we rely on touch either during encoding or retrieval. This would mean that a tactile retrieval cue is more effective than a visual retrieval cue in retrieving information which was presented visually in the encoding phase. Studies in cross-sensory object recognition and recall, which have presented a variety of visuotactile and audiovisual stimuli (including non-semantic stimuli), however, show that cross-sensory performance may come at a cost, but never at a benefit (Bulter & James, 2011; Ernst, Lange & Newell, 2007; Woods & Newell 2004) Furthermore, a cross-sensory interpretation is also at odds with the literature on the beneficial effects of reinstating encoding operations at retrieval (Dewhurst & Knott, 2010;

Kent & Lambers, 2008; Tulving & Thompson, 1973). According to that literature, a retrieval cue should be more effective when it is more similar to the encoding cue. Thus, the benefits of adding touch to vision are most likely not due to a cross-sensory mechanism. More likely, the multisensory benefits reported here are due to sensory interactions between the information processing in multiple sensory modalities (see Shams & Seitz, 2008) or are due to multiples sensory modalities independently initiating encoding and/or retrieval processes (i.e., along the lines of dual-coding; Paivio, 1971; 1986). Sensory interactions may result in enhanced encoding and retrieval processes (see Figure 5.1b) which increase memory performance. In the second case, it should be noted that cross- or supra-sensory may occur *in addition* to within-sensory encoding and retrieval processes. For instance, when a visual-tactile retrieval cue is used to retrieve a visual item, both a visual and a tactile retrieval attempt may be initiated. Even tough tactile retrieval of a visual item may not be as effective as visual retrieval of a visual item, if it occurs in addition to, it may increase memory performance.

Conclusions

To summarize, we have found benefits for adding touch to vision during encoding and retrieval of object identity and object location memory. These results show that the benefits of multisensory memory are not limited to audiovisual objects, to encoding information, and to object identity memory but are also present for visuotactile objects, for retrieving information, and for object location memory. However, it is yet unclear whether these multisensory benefits are the result of multisensory interactions or of multiple encoding and retrieval operations initiated by multiple sensory modalities.

Chapter 6: Multisensory memory effects differ for congruent semantic and non-semantic items

Abstract

Separate studies reported larger multisensory memory effects for semantic than for non-semantic items with different experimental paradigms. We study multisensory memory effects for both item categories with a similar paradigm. Sixty participants were sequentially presented with either auditory (A), visual (V), or audio-visual (AV) representations of either semantic items (familiar objects) or meaningless items (Morse codes). After item presentation, participants recalled as many items as they could from memory. For semantic items, we found that memory performance in the AV condition was significantly higher than in the V condition (30%; $p < .05$) and the A condition (60%; $p < .001$). For non-semantic items, we found no differences in memory performance between the conditions ($p = .054$). In a control experiment we showed that these findings are independent of whether we visually present the non-semantic items as spatial or as temporal patterns. Our findings suggest that the semantic information carried by the familiar objects plays an important role in multisensory memory benefits.

6.1 Introduction

Multisensory memory benefits

In the previous chapter we explored the favorable effects of visuo-tactile presentation on the encoding and retrieval of object recognition and object location memory. Participants were presented with four Morse codes at four locations in the visual (V), the tactile (T) or both the visual and the tactile modality (VT). Thereafter, they were again presented with eight Morse codes and had to indicate if and where they had experienced these codes earlier. For object identity memory, we found that visuo-tactile presentation improved memory encoding and retrieval by about 10% in comparison with visual-only presentation, but did not improve in comparison with tactile-only presentation.

In contrast with our results are those obtained earlier by Thompson and Paivio (1994). These authors serially presented twenty familiar objects either as pictures (V), sounds (A), or combinations of congruent picture-sound pairs (AV); and asked participants to recall as many items as they could. The results showed that the mean score in the AV condition was approximately 50% higher than in the A and the V condition. So, here multisensory memory performance was better than in either modality.

Possible causes of the different multisensory memory benefits

The different multisensory effects reported in these two studies (i.e., by Thompson and Paivio and in the previous chapter) can be caused by the different items that where presented. Thompson and Paivio presented sounds and pictures of familiar objects. These items were semantic, the pictures were presented as spatial patterns, and different (complementary) patterns (with the same meaning) were presented to each sense. In contrast, the items presented in the previous chapter did not carry semantic information (because the participants were unaware of their meaning). Also, the Morse codes were visually presented as temporal patterns, which were identical to the tactilely presented patterns (i.e. the patterns were redundant). The two studies also had several methodological differences which may have caused different multisensory memory effects. First, different modality combinations were investigated (viz., AV and VT). Second, participants were tested with either a free recall (Thompson & Paivio) or a recognition task (previous chapter). Third, in the study by Thompson and Paivio participants were automatically presented with the items (a passive method), whereas we instructed participants to actively explore the items in the previous chapter. The differences in item presentation and methodology are summarized in Table 6.1.

Differences in semantic information carried by the stimuli as a possible cause

In this chapter we investigate the different effects of the presented items (e.g. familiar objects and Morse codes) on multisensory memory. We expect that multisensory memory effects for familiar objects and Morse codes differ because familiar objects carry semantic information and Morse codes don't (at least they don't for people who are not familiar with them). This expectation is based on the earlier results obtained by several studies on the role of semantic information in multisensory memory processing (Laurenti, Kraft, Madljian, Burdette & Wallace, 2004; Lehmann & Murray, 2005).

Murray and colleagues (Murray *et al., 2004*; Lehmann & Murray, 2005) conducted several studies in which they continuously presented sounds and pictures of familiar objects. For each item participants had to indicate whether they had seen it before. The results show a small but significant advantage of audiovisual presentation over visual presentation. Importantly, in the Lehmann and Murray study, this advantage was only present for audiovisual items that consisted of a sound and a picture of the same familiar object. When the sound was replaced by a non-semantic 1000 Hz tone, or by a tactile 'tap' to the body, the advantage disappeared. This indicates that semantic information carried by the presented stimulus plays a role in multisensory memory processing.

The role of semantic information has also been highlighted by studies on multisensory object discrimination (Laurenti *et al*, 2004) and object identification (Suied, Bonneel & Viaud-Delmon, 2009; Yuvall-Greenberg & Deoull, 2007; 2009). These studies show that the semantic information carried by multisensory stimuli affects behavior. Yuvall-Greenberg and Deoull, for instance, presented sounds and pictures of animals and after the presentation of each animal they asked participants whether they had just experienced a specific animal in a specific modality. Presentations were either 'congruent' (same animal in both modalities), 'neutral' (animal with a non-semantic object), or 'incongruent' (different animals). They found that reaction speed increased for incongruent presentations, but decreased for congruent presentations, in comparison with neutral presentations.

The above results indicate that the different multisensory memory benefits reported by Thompson and Paivio and in chapter 5 may be due to differences in the semantic information carried by the items presented in each study. However, a definite conclusion cannot be drawn because in the studies mentioned in the previous two paragraphs, non-semantic items were presented differently than in chapter 5. Specifically, in those studies, non-semantic items presented in one modality (such as a 1000 Hz tone) were always paired with semantic items presented in another modality. This means that these items were semantically incongruent. In contrast, the Morse codes presented in the previous chapter were semantically congruent (see Table 6.2 for a summary of these differences). Thus, it is not yet clear how the multisensory memory benefits for congruent semantic items relate to those for congruent non-semantic items.

Research objective

The objective of this study is to investigate the multisensory memory effects for congruent semantic and congruent non-semantic items. We replicated Thompson and Paivio's experiment, both with the original items (i.e. familiar objects) and with the items from the previous chapter (i.e. Morse codes). We hypothesize that familiar objects and Morse codes lead to different multisensory memory effects. Specifically, for the familiar objects we expected to replicate the multisensory effects reported by Thompson and Paivio (i.e., AV better than A and V), and for the Morse codes we expected to replicate the multisensory effects in the previous chapter (i.e., AV not better than A or V).

Table 6.1: Differences in item presentation and methodology

Type	Difference	Thompson & Paivio	Previous chapter
Item	Semantics	Semantic	Non-semantic
	Visual presentation	Spatial pattern	Temporal pattern
	Multisensory presentation	Complementary	Redundant
Methodology	Modalities	Audition and vision	Touch and vision
	Task	Recall	Recognition
	Exploration	Active	Passive

This table details the differences in the presented items and metholodogy in the study by Thompson & Paivio (1994) and that of the previous chapter (i.e., chapter 5). The first two columns list the difference and the third and fourth list the respective values for the study by Thompson & Paivio and that of the previous chapter.

As shown in Table 6.1, the differences between Morse code and familiar objects are not limited to differences in semantics. The visual component of familiar objects is a spatial pattern whereas the visual component of Morse codes is a temporal pattern. The visual modality, however, is more appropriate to process spatial (e.g. textures) and spatiotemporal patterns (e.g. motion) than temporal patterns (see Baddeley, 1986), which also may affect multisensory memory effects. For instance, if memory for visual Morse codes is relatively poor they may not contribute to memory performance during the presentation of audiovisual Morse codes. To control for this hypothetical effect, we conducted a control experiment were visual Morse codes were presented as spatial instead of temporal patterns. Sensory components of Morse codes are also redundant whereas the sensory components of familiar objects complement each other. This difference is also controlled for in the control experiment. When visual Morse codes are presented as spatial instead of temporal patterns they are no longer redundant with the auditory temporal patterns.

Table 6.2: Properties of the items studied in multisensory memory studies

Multisensory memory study	Studied items carried semantic information?	Studied items were semantically congruent?
Thompson & Paivio; Lehmann & Murray	Yes	Yes
Lehmann & Murray	Partially	No
Chapter 5	No	No

This table details the semantic properties of the items presented in several multisensory memory study. Note that Lehmann & Murray (2005) as well as several other studies (see text) compared multisensory memory effects for congruent semantic items with items that had different properties than those presented in chapter 5.

6.2 Method

Participants

Sixty volunteers from TNO (39 men, 21 women, mean age = 38.2 years, SD = 10.7, age range: 22–62 years) participated in the experiment. All participants reported normal or corrected-to-normal vision, normal hearing, and were without other cognitive abnormalities. Furthermore, participants reported having no previous experience with Morse codes and were naïve to the purpose of the present study. Prior to inclusion, all participants gave their written informed consent. Data from three participants were excluded from the analyses due to illness or withdrawal and replaced with data from three new volunteers.

Setup

Participants were seated in front of a 17 inch CRT touch screen and wore Sennheiser headphones (HD 465). The screen was produced by Elo and displayed an area of 800 × 600 pixels with a frequency of 85 Hz. The distance between the screen and the seated participant was approximately 55 cm.

Stimuli

Items were 26 familiar objects and 20 Morse codes. These could be presented auditorily, visually, or audio-visually. Familiar objects were selected based on Thomspon and Paivio's early study (1994) on multisensory memory benefits. They consisted of natural objects familiar from everyday life such as a guitar, a cat, a bird, scissors or a fly. Visual presentations of the familiar objects were line drawings of easily identifiable objects selected from Snodgrass and Vanderwart (1980). The line drawings employed in the current study are presented in the appendix. Corresponding auditory presentations were collected from various sound-databases on the world-wide web. Audio-visual presentations consisted of a simultaneous presentation of both a line drawing and its corresponding sound. Each line drawing was presented at the centre of the screen.

Morse codes were similar to those presented in the previous chapter. Each Morse code consisted of one to four dots and/or dashes. The duration of the dots and dashes were 200 ms and 600 ms, respectively, separated by intervals of 200 ms. Visual Morse codes were presented at the centre of the screen as a flashing white disc (Ø = 2 cm) on a gray background. Auditory Morse codes were presented as 100 Hz tones. Audio-visual Morse codes consisted of a simultaneous presentation of a visual and auditory code. The complete list of all stimuli employed in this study is also shown in the appendix.

From these stimuli, six different lists were constructed: a list of familiar objects and a list of Morse codes for each of the three sensory conditions (i.e,. A, V, and AV). The order of the items on the lists was the same for A, V, and AV.

Design and procedure

Overall, we adopted the setup and procedures from Thompson and Paivio (1994). Participants were divided into three equal sized sensory condition groups (i.e. A, V, or AV) with 20 participants in each group. The groups were balanced for age and gender. Participants were tested individually in a silent room and took part in either two auditory, or two visual or two audio-visual conditions (i.e., one with familiar objects and one with Morse codes). The order in which participants received both conditions was counterbalanced: i.e., 10 started with the familiar objects, the other 10 started with the Morse codes. Additionally, an interval of at least one hour between conditions was employed.

For all conditions, memory performance was assessed by means of a free recall task adapted from Thompson and Paivio. Each condition consisted of the presentation of either unisensory (i.e. A or V) or multisensory (i.e. AV) item sequences. Visual familiar objects were presented for 1.1 s and corresponding auditory objects were presented for 3.5 s. As Thompson and Paivio (1994) point out, “This discrepancy […] reflects the fact that the sounds require more time for accurate identification than the pictures do.” (p. 384). The onset of the visual (1.1s) and auditory (3.5 s) components of the audio-visual item was simultaneous and accordingly the sound outlasted the image. In all the above conditions, the presentation of each familiar object was separated with an inter-stimulus interval (ISI) of 0.9 s. Morse codes, which had durations varying from 0.2s (one dot) to 3.0 s (four dashes) all had a stimulus onset asynchrony (SOA) of 4.4 s (i.e., equal to the SOA in the auditory and audiovisual familiar object conditions).

Participants were instructed prior to the presentation of each list that they would be tested for object memory. They were also required to perform a rehearsal-inhibiting distracter task: during item presentation participants had to count backwards by threes from the number 200 as quickly as possible. Similar to the methods of Thompson and Paivio (1994), they had to record the count by writing down the intermediary sums on a sheet of paper while the items were presented. Thus, participants had to count and write while they listened to and/or watched the items. They were instructed to perform these two tasks simultaneously to the best of their ability. To ensure equal difficulty across visual and auditory conditions, participants had to look up at the screen while they wrote in all conditions, even those who were not presented with visual items. So, participants had to write down the sums without looking at their sheet. Given the difficulty of the dual-task, participants were given approximately 30 s to practice counting backwards, prior to the presentation of the item list.

After presentation of the items, participants were given another sheet of paper and were asked to write down the names of as many familiar objects or Morse codes as they could recall from memory. For Morse codes, participants were instructed to write them down as dots and dashes. For instance, a code consisting of a dot followed by two dashes should be written down as: ● ▬ ▬. For this, all participants were given as much time as they required. Finally, again following the procedure laid down by Thompson and Paivio, participants again received a same list with familiar objects and were asked to name all individual items sequentially. The labels were utilized to score participants’ recall responses in such a way that the label most frequently reported for a particular familiar object in each condition was considered to be the only correct answer.

Control experiment
In the control experiment, the visual and audiovisual Morse code lists were presented again but with visual Morse codes presented as a single picture instead of a temporal pattern.
The control experiment was conducted approximately 5 months after the main experiment. It differed from the main experiment in the following ways. Twelve of the forty original participants from the visual and audio-visual groups could not participate again and were replaced with twelve different volunteers from TNO with approximately the same age and with the same gender. To ensure that the twelve new volunteers had about equal experience with the paradigm as the original participants, they were subjected to a pilot with a list containing familiar objects before they were admitted to the control experiment.
Only two conditions were tested in the control experiment, viz. a visual (V) and an auditory-visual (AV) condition. In both conditions only Morse codes were presented. The Morse codes were the same as in the main experiment, but the visual Morse codes were rendered as bitmap images instead of flashing discs. The dots (Ø = 2 cm) and dashes (2 cm [Height] × 6 cm [Width] rectangles) were now black against a white background. Just like the visual familiar objects in the main experiment, the visual Morse codes were displayed for 1.1 s, with an ISI of 0.9 s in the visual condition. In the audio-visual condition, the auditory Morse code often outlasted the visual Morse code, although the SOA remained 4.4 s.

6.3 Results

After the experiments, all participants were asked about the purpose of the study. All stated that the study was conducted to assess general memory performance in the presence of an attentional distracter. Also, all participants declared that the attention to the counting task and the memory task was divided approximately 50/50, as per the instruction to perform both tasks simultaneously.

Main experiment
Figure 6.1 displays memory performance in all sensory conditions of the main experiment for both familiar objects (left panel) and Morse codes (right panel). Performance for the familiar objects is approximately twice as high as for the Morse codes. Two separate ANOVAs were carried out on the object data and on the Morse code data, each with Sensory Condition (three levels, i.e. visual, auditory, audio-visual) as between subjects variable. For the familiar objects, we found a significant main effect of Sensory Condition ($F_{2,57} = 9.03$; $p < .001$). A subsequent Tukey HSD post-hoc analysis revealed that the audio-visual condition differed significantly from each of the unisensory conditions (V [$p < .05$] and A [$p < .001$]). The unisensory conditions did not differ from each other ($p = .24$). For Morse codes, we found no significant main effect of Sensory Condition ($F_{2,57} = 3.07$; $p = .054$). Although this analysis approaches significance, the mean scores in the right panel of Figure 6.1 show that there is no indication that the audiovisual condition leads to a larger number of recalled items than both the visual and the auditory condition.

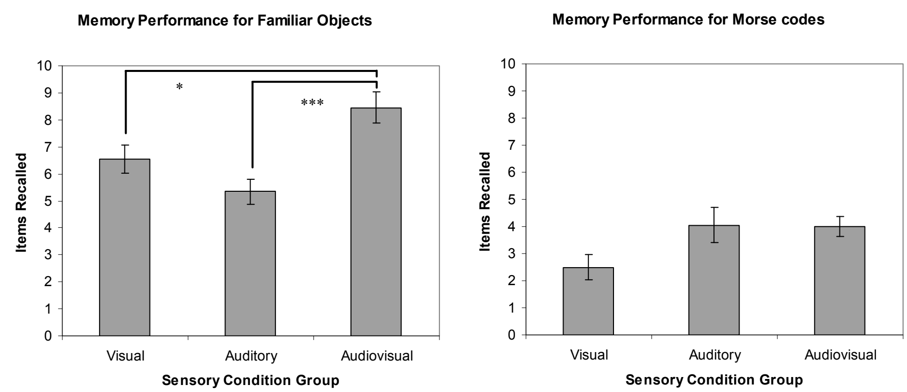


Figure 6.1. Mean number of correctly recalled items in each sensory condition for familiar objects (left panel) and Morse codes (right panel). In the left panel, the audio-visual condition differed significantly from both visual and auditory conditions. In the right panel, there were no significant differences. Significant differences are denoted by asterisks: one asterisk denotes a p < .05- level difference and three asterisks a p < .001-level. Error bars represent the standard errors.

Control experiment

The visual and audiovisual conditions measured in the control experiment are displayed together with the auditory condition from the main experiment in Figure 6.2. In the control experiment, the mean number of items recalled was 4.5 and 5.1, respectively, for the visual and audiovisual condition, as opposed to scores of 2.5 and 4.0 for the same conditions in the main experiment. An ANOVA on Sensory Condition (3 levels) revealed no significant main effect ($F_{2,57} = 1.19$; $p = .31$).

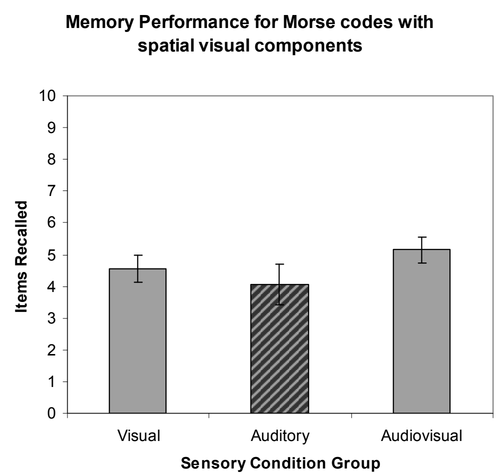


Figure 6.2. Mean number of correctly recalled items in each sensory condition for Morse codes with spatial visual components. The visual and audiovisual conditions were measured in the control experiment; they were compared to the auditory condition from the main experiment. There were no significant differences. Error bars represent standard errors.

6.4 Discussion

We investigated the multisensory memory effects for familiar objects and Morse codes with a free recall task. Because previous findings indicate that multisensory memory benefits are larger for congruent stimuli carrying semantic information, we hypothesized that multisensory presentation of familiar objects would enhance memory performance when compared to unisensory presentation. In contrast, no such effect of multisensory presentation was anticipated with regard to Morse codes. The current results confirm this hypothesis.

The data of the main experiment show that the mean performance for familiar objects is approximately twice as high relative to the performance for Morse codes. This means that familiar objects are easier to remember than Morse codes. The memory benefits of multisensory presentation are also larger for objects than for Morse codes (which are actually non-existent in this study). This means that familiar objects and Morse codes cause different multisensory memory benefits.

The mean scores for familiar objects are consistent with those reported by Thompson and Paivio (1994). They reported that about six items were recalled in the unisensory conditions, and about nine items in the multisensory condition. In our experiment, we found that the number of items recalled in the unisensory conditions was between 5-7 and about 8.5 in the multisensory condition. For Morse codes, the number of items recalled in the main experiment is in line with the results reported in the previous chapter. In both studies the number of Morse codes memorized was about 2.5 to 4. Thus, we successfully replicated the results of Thompson and Paivio's experiment, as well as those from the previous chapter.

Why do multisensory memory benefits differ for familiar objects and Morse code?

Now that we have established that multisensory memory benefits are larger for familiar objects than for Morse codes we consider which mechanism caused this effect. In the introduction we discussed three differences between familiar objects and Morse codes. First, familiar objects carried semantic information, while Morse codes didn't (because our participants were not familiar with them). When we process information with multiple sensory modalities it is checked for congruency. When information is congruent multisensory interactions occur. Beneficial sensory interactions may only occur at the memory level when semantic information from multiple sensory modalities is semantically congruent. Second, the visual component of familiar objects was a spatial pattern while the visual component of Morse code (in the main experiment) was a temporal pattern. Generally, the visual modality is more appropriate to process spatial or spatiotemporal patterns, but less suited for temporal patterns (Baddely, 1986). While visual Morse codes patterns presented in this experiment were perfectly perceivable (see chapter 5's control experiment), they may not have been processed properly in memory and may not have contributed to the performance in the audiovisual Morse code condition. Third, multisensory memory benefits may be differ when unisensory components are complementary instead of redundant. When multisensory presentation is complementary, more information is presented. This may increase the chance that some information is remembered (e.g. complementary information may be processed by more systems than redundant information), increasing memory performance.

To distinguish between these mechanisms we conducted a control experiment. In that experiment, we presented visual Morse code as spatial instead of visual patterns. As a result of this change the unisensory components of the Morse code presented in the audiovisual condition were also complementary instead of redundant. However, even in the control experiment we still did not find any significant multisensory memory benefits for Morse codes. The absence of a multisensory memory effect for Morse code in both the main and the control experiment can only be explained by the difference that the semantic information carried by the stimulus is a critical factor for multisensory memory benefits.

On multisensory memory benefits for semantic stimuli

The role the semantic information plays in multisensory memory benefits has been investigated earlier in several studies (Doehrmann & Naumer, 2008; Laurenti, Kraft, Madljian, Burdette & Wallace, 2004; Yuvall-Greenberg & Deouell, 2007; 2009). In those studies, however, multisensory memory benefits were found when sounds and pictures were semantically congruent, but not when a sound or picture (i.e. a semantic item) was presented in one sensory modality together with a non-semantic item in another sensory modality. The results of this study complement these results by showing that multisensory memory benefits also differ between congruent semantic items and congruent non-semantic items (i.e. presenting two non-semantic items referring to the same object). Thus, our results further highlight the role semantics play in processing information from multiple sensory modalities.

Conclusions

To summarize, we investigated if the semantic information carried by familiar objects and Morse codes led to different multisensory memory benefits. In the main experiment, we found that multisensory memory performance was better than unisensory memory performance for familiar objects but not for Morse codes. This shows that differences in stimulus characteristics of familiar objects and Morse codes caused different multisensory memory benefits. In a control experiment, we altered the visual component of the Morse codes and with it, the redundancy between the two sensory components. The results of that experiment show no significant enhancement in multisensory memory performance. Taken together with the results of the main experiment, this indicates that the semantic information carried by familiar objects leads to multisensory memory benefits.

Chapter 7: Audiovisual navigation in virtual mazes

Abstract

In comparison to unisensory presentation, multisensory presentation can improve perception, attention, and memory. We investigated the effects of multisensory presentation on spatial memory and navigation in virtual environments. We hypothesized that multisensory presentation improves both spatial memory and navigation. Nineteen participants explored three virtual mazes consisting of 10 nodes and 13 corridors. Each maze contained nodes either with visual, auditory or audiovisual landmarks. Each maze was explored for 90 seconds. After each exploration, participants performed the following tasks in fixed order: 1) draw a map of the maze, 2) recall adjacent landmarks for three given landmarks, 3) place all landmarks on a map of the maze, and 4) find their way through the maze to locate five given landmarks in fixed order. Significant improvements after exploration with multisensory landmarks over exploration with unisensory landmarks were present in the maze drawing task ($p < .001$), the adjacency task ($p < .01$), and the wayfinding task ($p < .05$). We observed no differences following unisensory and multisensory landmark presentation in the landmark placement task ($p = .66$). We conclude that in comparison to unisensory presentation, multisensory presentation improves spatial memory and navigation performance in virtual environments.

7.1 Introduction

In the previous two chapters we reported that multisensory presentation improves object identity and object location memory. In this chapter we investigate the effect of multisensory presentation on spatial memory and navigation in virtual environments. Navigation is the process of planning and following routes using perception and spatial memory to travel from the current location to a goal. It is indispensible for finding one's way in an environment and therefore essential in our daily lives.

In the acquisition of spatial memory, landmarks play an important role (Siegel & White, 1975; Montello, 1998). Landmarks are typically distinctive objects that stand out in the environment (Presson & Montello, 1988; Caduff & Timp, 2006). They serve as reference points and are therefore extremely helpful when we are following routes or when we need to determine where we are (Janzen, 2006).

In real world navigation, the components of a landmark may be visual, but could also consist of information received in other sensory modalities. It stands to reason that these non-visual components may aid us during navigation. For instance, the sound of traffic could inform us that we are close to a highway. Likewise, the smell of fish or salt may tell us that we are close to the fish auction or to the sea. But what happens when we have both visual and non-visual information available at the same time?

Previous research on the simultaneous processing of information with multiple sensory modalities indicates that it can be advantageous (Ernst & Bülthoff, 2004; Philippi, Van Erp & Werkhoven, 2008; Stein & Meredith, 1993; Werkhoven, van Erp & Philippi, 2009). Studies in perception show that, in comparison to unisensory presentation, multisensory presentation enhances reaction time (Bernstein, Clark & Edelstein, 1969; Hershenson, 1962; Nickerson, 1973) and reliability (Ernst & Banks, 2002; Ernst & Bülthoff, 2004; Shams, Ma & Beierholm, 2005). Likewise, in memory research, it has recently been shown that multisensory experiences enhance recall and recognition in object identity (Thompson & Paivio, 1994; Delogu, Raffone & Olivetti Belardinelli, 2009) and object location (see chapter 5).

In the first chapter of this thesis, we introduced the Trisensory Orientation Model (TOM) to model the effects of congruent multisensory presentation on navigation. The model assumes a sequence of perceptual, memory and navigation processes. According the model, multisensory enhancements in perception and (spatial) memory should affect navigation. In this chapter, we therefore investigate whether the presentation of multisensory landmarks in a virtual environment improve spatial memory and navigation performance.

Earlier work on audiovisual benefits in virtual environments

When navigating virtual environments, landmark information is often limited to the visual sense. Two studies have investigated the effect of audiovisual presentation on navigation performance. Gunther, Kazman, & MacGregor (2004) investigated navigation in virtual environments that contained visual objects which could produce 3D sound. They found that the addition of 3D sound to the virtual environments improved navigation, but did not increase a participant's spatial memory of the environment. However, the authors explained that the sound was audible even when participants did not look at the objects (i.e., it was audible through the walls) and that this worked as a beacon to guide participants to their destination. Because the sound was often observed without the corresponding visual object, multisensory interactions may have played no role in this study.

Another study on the effects of multisensory presentation in virtual environments was conducted by Ardito, Costabile, De Angeli, and Pittarello (2007). These authors played different classical music in each room of a virtual museum. They found that this music could benefit users' navigation and memory performance, but only when users where informed in advance of the link between the music and the rooms. The authors suggested that for stimuli with a 'natural' link (viz. semantic stimuli) the benefit on navigation and memory performance may be automatic. This suggestion is in agreement with the recent studies (including our work in the previous chapter) which indicate that meaning plays a critical role in multisensory memory interactions (Laurenti, Kraft, Maldjian, Burdette & Wallace, 2004; Yuvall-Greenberg & Deouell, 2007; 2009)

The effect of meaningful multisensory landmarks in navigation

The first objective of this study is to explore the effects of multisensory presentation on spatial memory in virtual environments. The second objective is to investigate whether these effects extend to navigation. To investigate these objectives we constructed several virtual mazes in the computer game *Unreal Tournament 2004.* The mazes either contained auditory, visual or audiovisual landmarks. The landmarks were sounds and pictures of meaningful semantic objects. We conducted an experiment to compare the effect of the landmark presentation modality on the user's spatial memory and navigation performance. Our first hypothesis, based upon earlier findings in chapter five, is that multisensory landmark presentation improves spatial memory of the virtual mazes. Our second hypothesis, which is based upon our model, is that it also improves navigation performance.

7.2 Method

Participants

Nineteen (10 male, 9 female) students with ages between 19 and 28 (mean age = 22.2; SD = 2.14) participated in the experiment. All had normal or corrected-to-normal vision and normal hearing. Participants rated their own spatial and memory abilities on a 5-point scale at 3.4 (SD = 0.8) and 3.6 (SD = 0.7), respectively. Ten participants did not regularly play any three-dimensional computer games. The other participants either played such games monthly (3), weekly (5) or daily (2). All participants gave their informed consent prior to participation and completed the experiment in approximately 50 minutes. Participants were paid for their participation.

Experimental setup

Participants were seated approximately 70 cm from an Iiyama 24 inch LCD monitor. The monitor displayed a virtual maze with a resolution of 1680 (H) by 1050 (V) pixels. Navigation through the virtual maze was by keyboard, which lay in front of the monitor.

Participants could press the arrow keys to move forward or backward, and to turn left or right. When participants navigated the virtual maze they wore headphones (Sennheiser HD150).

The virtual mazes were rendered in *Unreal Tournament 2004* (Atari, New York City). Participants had a field of view of approximately 90 (H) by 56 (V) degrees which was located about 1.75 m (or 88 Unreal Units [UU]) above ground-level. Movement speed was roughly 9 m/s (or 480 UU/s) and rotation speed was 150 degrees per second.

For this experiment, four different mazes were constructed with the level editor of *Unreal Tournament 2004*. Each maze consisted of 10 nodes connected by 13 corridors. The nodes were standing cylinders with a radius of 10 m (512 UU) and a height of 5 m (256 UU). A white cube (2.5 m; 128 UU) was positioned in the center of each node. These cubes were used for landmark presentation. The corridors were bars with a width and height of 5 m (256 UU) and lengths varying between 16 m (800 UU) and 60 m (3000 UU). The corridors were placed in such a way to prevent participants observing multiple cubes in a single screen. A topside view of each maze is presented in Figure 7.1; An impression of a maze is shown in Figure 7.2.

Stimuli

Each maze contained ten landmarks which were presented either auditorilly, visually or audiovisually at or from the white cubes located in the center of each node. The identity of each landmark presented was fixed (see Table 7.1).

Visual landmarks were line drawings selected from the Snoddgrass and Vanderwaart (1980) set. They were presented on the sides of the white cubes (see Figure 7.2 for a list of the items and the appendix for the images). Auditory landmarks were sounds (44Khz, 66 dB[A]) matching the identity of the line drawings. The sounds were gathered from the internet and were all modified to a duration of 3.5 seconds. *Unreal Tournament 2004* automatically modulated the stereophonic components of the sounds to create the illusion that the sounds emanated from the white cubes (which were present with blank sides when used for auditory only landmark presentation). Sounds started playing when the corresponding white cube was present in the participant’s field of view. Once started, a sound continued playing until the end of its duration. When a white cube was still (or again) in a participant’s field of view after a sound finished playing, that sound was played again after an interval of 0.5 seconds.

Audiovisual landmarks were line drawings and matching sounds presented simultaneously. That is, during audiovisual landmark presentation the white cubes had line drawings on their sides and played a matching sound when they were inside the participant’s field of view.

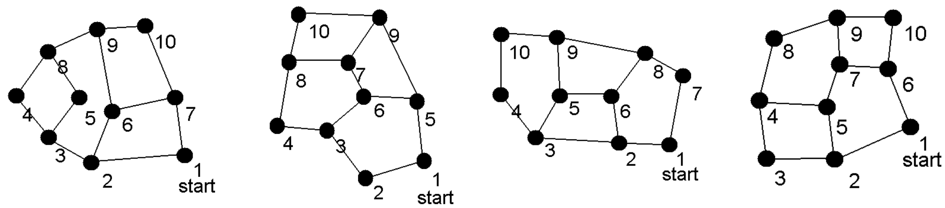


Figure 7.1: Top-down view of the four mazes. From left to right: maze 1, maze 2, maze 3, and maze 4. Each maze had 10 nodes and 13 corridors. The nodes are numbered 1 to 10.

Design

The experiment consisted of one training and three experimental conditions. In each condition, participants had to explore one of the four mazes and were asked to complete four tasks. All participants started with maze 1, which was used solely to familiarize participants with the experimental procedure. Maze 1 always contained three visual, three auditory and four audiovisual landmarks (see Table 7.1).

Thereafter, Mazes 2 to 4 were completed in fixed order. Each of these mazes either contained ten visual, ten auditory, or ten audiovisual landmarks. Each participant completed one maze containing only visual landmarks (the visual condition), one maze containing only auditory landmarks (the auditory condition) and one maze containing audiovisual landmarks (the audiovisual condition). The order of which condition was presented in which maze was balanced across participants.

Figure 7.2: A screenshot from the virtual maze. In this instance, the user is at the first node in the third maze. The white cube located at this node shows the line drawing of a donkey. If the user was exploring this maze and it contained audiovisual landmarks then he or she would also hear the donkey bray when he or she looked at the cube.

Table 7.1 The landmarks presented in each maze

Node	Landmarks Maze 1	Landmarks Maze 2	Landmarks Maze 3	Landmarks Maze 4
1	Cannon (AV)	Whistle	Donkey	Bird
2	Camera (A)	Saw	Train	Pistol
3	Church (AV)	Fly	Scissors	Helicopter
4	Tree (V)	Pencil	Frog	Cat
5	Umbrella (V)	Telephone	Apple	Bell
6	Pig (AV)	Clock	Bicycle	Accordion
7	Rubber Duck (A)	Cow	Airplane	Horse
8	Music Box (A)	Drum	Trumpet	Piano
9	Ball (AV)	Car	Guitar	Seal
10	Grasshopper (V)	Duck	Chicken	Toothbrush

This table lists the landmarks presented at each node (rows) in the four mazes (columns). The node numbers correspond to those in Figure 7.1. Of the landmarks in the first maze, three were always presented visually (V), three auditorilly (A) and four audiovisually (AV). The landmarks in the other mazes could either be presented visually, auditorily or audiovisually.

Procedure and tasks

At the start of the experiment, participants were given written instructions explaining the experimental procedure. In each maze, participants were inserted at the first node (see Figure 7.1) and had 90 seconds to explore that maze. For comparison, in each maze it took the experimenter – who was familiar navigating these mazes – about 45 seconds to visit each node and about 70 seconds to walk through each corridor at least once. The exploration time was set at 90 seconds because pilot studies indicated that when exploration time was longer than 90 seconds participants often achieved perfect scores in some of the subsequent spatial memory tasks and when exploration time was less than 90 seconds participants often did not manage to visit all the nodes in the maze.

Participants were told in advance in which modalities the landmarks would be presented in that maze and were encouraged to visit each node in the maze at least once. During the exploration of each maze, participants also had to verbally repeat the letter string ‘a-b-c’. This articulatory suppression task was used to prevent verbal recoding of the spatial features of the maze and the identity of the landmarks (see Thompson & Paivio, 1994).

After each exploration participants had to complete four tasks in fixed order. This order was chosen because in any other order participants would have to report information that was presented in a previous task. Participants had a fixed amount of time to complete each of the first three tasks. When one of these three tasks was completed early, participants had to wait until the time for the task expired. The tasks were adopted from earlier studies exploring spatial memory and navigation (Darken & Sibert, 1996; Goldin & Thorndyke, 1982; Rovine & Weisman, 1989). They were selected because pilot studies indicated that performance on these tasks was affected by landmark presentation modality. The first three tasks were used to assess spatial memory, while navigation performance was measured with the fourth task.

Task 1: Maze drawing task

In the first task participants were given an empty sheet of paper and asked to draw a map of the maze they had just explored. In addition to drawing the nodes and the connections they had to indicate which landmark was present at which node by writing down the name of that landmark near each node. Participants had 90 seconds to complete this task.

Task 2: landmark adjacency task

In the second task participants were given a sheet of paper containing a list of three landmarks from the maze they had explored. For each landmark, they had to write down all the landmarks which were directly connected with it (i.e. through a single corridor). For example, the landmark *tree* is directly connected with the landmarks *music box* and *church* (see Figure 1 and Table 1). The landmarks on the lists were not directly connected with each other. They were *tree*, *camera*, and *pig* for maze 1; *saw*, *cow*, and *telephone*, for maze 2; *train*, *apple*, and *frog* for maze 3; and *bell*, *accordion*, and *piano* for maze 4. Participants were given 60 seconds to complete this task.

Task 3: landmark placement task

In the third task participants were given a sheet of paper with the actual top-down view of the maze (as in Figure 7.1) and a list of all the landmarks encountered in that maze. Here, they had to place each landmark on the correct node of the maze. Participants had 90 seconds for this task.

Task 4: wayfinding task

In the last task participants had to find their way in the virtual maze. They were inserted in the maze and had to navigate to another node with the shortest possible route in the fastest possible time. On insertion, the experiment leader announced to which node the participant had to go. Once the participant arrived at the appropriate node, the procedure was repeated until the participant had located a total of five nodes. In each maze, participants were dropped opposite of the starting node of the exploration phase. Navigating to the target node required them to traverse at least one or two other nodes. The paths they had to walk in each maze all had about similar patterns and distances. The landmarks they visited were, in fixed order: *tree* (starting point), *umbrella*, *grasshopper, cannon, scissors*, and *ball* for maze 1; *pencil* (starting point), *duck*, *clock, whistle, fly,* and *car* for maze 2; *chicken* (starting point) *scissors*, *trumpet, donkey, bicycle* and *frog* for maze 3; and *piano* (starting point), *helicopter*, *horse*, *bird*, *toothbrush,* and *cat* for maze 4. There was no time limit for this task.

7.3 Results

Task analyses
The effect of the experimental condition on the performance during exploration and in all four tasks was investigated by employing separate repeated measures ANOVA's on the data for each task. Post-hoc Fisher LSD tests were conducted where necessary. The data acquired in the training condition were not analyzed.

Exploration phase
Each landmark was observed (i.e. looked at and/or heard) at least once in nearly all explorations. The only exception was that in one audiovisual exploration one participant did encounter only nine out of ten landmarks. In the subsequent memory and navigation tasks that participant performed above average.

On average, participants perceived a landmark 23.2 times (SD = 2.8) during each exploration. In 3.3 of these times (SD = 1.9) participants saw and/or heard the same landmark twice without seeing and/or hearing another landmark. Furthermore, they visited a landmark (i.e. they entered the cylindrical node area of each landmark) 15.0 times (SD = 2.2) during each exploration. So, when a participant saw and/or heard a landmark, he or she did not always move into that landmark's node area.

The number of landmarks observed, the number of times a participant observed a landmark twice in a row, and the number of times they visited a node did not differ between experimental conditions ($F_{2,36} = 1.3$; $p = .28$, $F_{2,36} = 0.2$; $p = .82$, and $F_{2,36} = 2.79$; $p = .07$, respectively).

Task 1: maze drawing task
The maze drawing task was scored for the number of recalled landmarks and for the number of drawn corridors between the nodes of the recalled landmarks. When a landmark's name was written somewhere on the map it was considered to be correctly recalled. When a corridor connected two recalled landmarks that were also connected by a corridor in the maze that was just explored it was considered to be correctly drawn. In other words, the relative location of the recalled landmarks and the drawn corridors was not taken into account for scoring. No penalties were applied when a landmark written and corridors drawn that where not present in the maze that was just explored.

The scores for the three experimental conditions are displayed in Figure 7.3. We found a significant effect of Experimental Condition (3) on the number of recalled landmarks ($F_{2,36} = 6.4$; $p < .001$) and on the number of drawn corridors ($F_{2,36} = 8.2$; $p < .001$). Both the number of recalled landmarks and the number of drawn corridors differed between the audiovisual (AV) condition and the auditory (A) and visual (V) conditions (all at least $p < .01$), but not between the A and the V condition ($p = .83$ and $p = .94$ respectively).

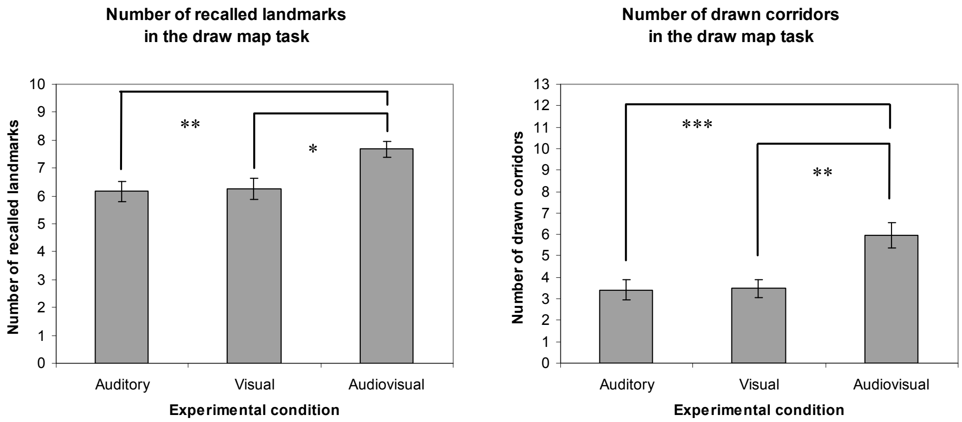


Figure 7.3: Performance in the maze drawing task as a function of experimental condition. Each panel displays performance for the auditory, visual, and audiovisual conditions. In the left panel, performance was scored for the number of recalled landmarks. In the right panel, performance was scored for the number of drawn corridors. Significant differences are indicated by asterisks: two asterisks denote a significance level of $p < .01$ and three asterisks of $p < .001$.

Task 2: landmark adjacency task

The numbers of recalled adjacent landmarks for each condition are displayed in Figure 7.4. No penalties were applied for incorrect responses. We found a significant effect of Experimental Condition (3) ($F_{2,36} = 4.2$; $p < .05$). The number of adjacent landmarks differed only for the AV and A ($p < .01$) condition and not for the AV and V ($p = .13$) and the A and V conditions ($p = .19$).

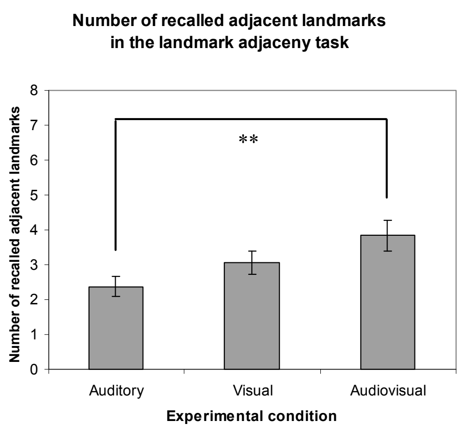


Figure 7.4: Performance for the landmark adjacency task as a function of experimental condition. Two asterisks denote significant differences of $p < .01$.

Task 3: landmark placement task

Participants correctly placed 4.78 (SD = 0.39), 5.42 (SD = 0.50), and 5.05 landmarks (SD = 0.56) on the correct node in the A, V, and AV conditions, respectively. There was no significant effect of Experimental Condition ($F_{2,36} = 0.4$; $p = .66$).

Task 4: wayfinding task

In the navigation task participants had to find their way to five landmarks in the maze. The optimal route to visit those five landmarks required a participant to pass (or visit) twelve nodes. With a map, it took the experimenter approximately 60 s to navigate this route. For one or more experimental conditions, three participants required more than the average time plus three times the standard deviation to find their way to the five landmarks. These participants were considered outliers and all data from these three participants were excluded from this analysis. The time it took the remainder of the participants and the number of nodes those participants visited to find the five target landmarks is displayed in Figure 7.5. We found a significant effect of Experimental Condition (3) for the number of visited landmarks ($F_{2,30} = 7.5$; $p < .01$) and the time taken ($F_{2,30} = 9.1$; $p < .001$). For the number of visited landmarks, AV differed from A ($p < .01$) and V ($p < .05$), but A did not differ from V ($p = .09$). For the time taken, AV differed from A ($p <.001$) and V ($p <.01$), but A did not differ from V ($p = .11$).
A correlation analysis between the average time taken and the average number of visited landmarks per participant revealed a strong correlation between these measures (*$r = 0.91$; $p < .001$*).

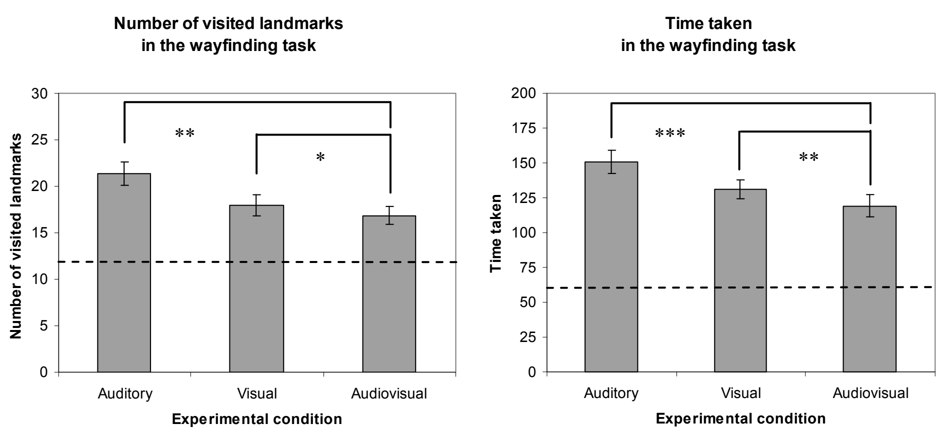


Figure 7.5: Wayfinding performance as a function of experimental condition. Performance was scored for the number of landmarks visited (left panel) and the time taken (right panel) to find all the five targets. Significant differences are indicated by asterisks: a single asterisk denotes significance level of *p* < .05, two asterisks denote a significance level of *p* < .01, and three asterisks of *p* < .001. The dashed lines indicate optimal performance.

7.4 Discussion

We investigated possible benefits of presenting multisensory landmark information in virtual environments. Specifically, we examined whether exploration of a virtual maze containing audiovisual landmarks improved spatial memory and navigation performance in comparison with exploration of a maze containing either auditory or visual landmarks. Spatial memory performance was measured with a maze drawing task, an adjacency task and a landmark placement task while navigation performance was measured with a wayfinding task. In two of the three spatial memory tasks and in the wayfinding task we find significant benefits after exploration with audiovisual instead of visual or auditory landmarks. These results show that audiovisual landmark presentation improves spatial memory and navigation performance with respect to both visual and auditory landmark presentation.

Multisensory benefits on spatial memory performance

The first objective of this study was to assess whether exploration with audiovisual landmarks improves user's spatial memory in comparison to exploration with either auditory or visual landmarks. Spatial memory was measured in three different tasks. The results for these tasks and their implications for the first objective are discussed below.

Spatial memory performance in the maze drawing task

The maze drawing task was scored for the number of recalled landmarks and for the number of drawn corridors. The number of recalled landmarks in the maze drawing task is in agreement with the number of recalled objects reported in the previous chapter and by Thompson and Paivio (1994). In both these studies sounds, pictures or sound-picture pairs were presented and it was found that participants recalled about six sounds, six pictures or nine sound-picture pairs. In our study, these numbers were about six, six and eight, respectively. The minor difference in the audiovisual scores may be explained by a ceiling effect: we presented only ten landmarks whereas the number of sound, pictures, or sound-picture pairs presented in the other studies was twenty or more.

The number of drawn corridors was low. On average, participants drew only 4 out of 13 corridors on their map. The number of drawn corridors in the audiovisual condition, however, did increase by approximately 70% in comparison with the visual and auditory conditions. This increase confirms our hypothesis that multisensory presentation can improve spatial memory in virtual environments and can be considered an extension of the multisensory benefits for object location memory we reported in chapter 5.

Interestingly, the multisensory benefit for the number of drawn corridors is more than twice as large as the multisensory benefit for the number of recalled landmarks. This is especially surprising considering that only the landmarks, but not the corridors, were presented multisensorially. This asymmetrical effect is likely related to the ceiling effect in the number of recalled landmarks. Because the number of recalled landmarks was close to a ceiling, it was theoretically impossible to obtain a multisensory improvement equal in size to the multisensory improvement in the number of drawn corridors. Furthermore, when participants have encoded all identities of the landmarks in the maze, they may have allocated (cognitive) resources that were involved in encoding landmark identities to encoding corridors, which would boost the multisensory memory benefit for the number of correctly drawn corridors even further.

Spatial memory performance in the adjacency task

In the second task, participants recalled more adjacent landmarks when they had explored a maze containing audiovisual landmarks than when they had explored a maze containing just auditory landmarks. However, in contrast with the results for the maze drawing task, we do not find a significant difference between the visual and audiovisual condition. This is a surprising result considering that one would expect that if a participant draws more corridors in the maze drawing task he or she should also be able to recall more adjacent landmarks. It should be pointed out, though, that the performance difference between the visual and audiovisual conditions approaches significance. The absence of a significant effect may be explained by decay of memory. A proportional decay of memory (Buhusi & Meck, 2006) decreases absolute performance differences over time. If noise remains constant subsequent memory tests will have less statistical power. In addition, participants were also asked about less landmarks in the adjacency task than in the maze drawing task. Obviously, the adjacency task had less statistical power than a similar task where adjacency for all landmarks would have been tested.

Spatial memory performance in the landmark placement task

In the third task, it did not matter whether participants had explored a world containing auditory, visual or audiovisual landmarks. In other words, we did not find a multisensory benefit. The absence of a benefit may also be the result of proportional decay. An additional, but (more) speculative explanation is that the alignment of the internal representation participants constructed during exploration of the mazes may not have matched the alignment of the map they were provided with (see Shelton & McNamara, 1999). Alignment mismatch affects can considerably affect performance (Palij, Levine, & Kahan, 1984; Presson, DeLange, & Hazelrigg, 1989; Richardson, Montello & Hegarty, 1999). Matching the internal representation to the map may have required complex operations (such as mental operations) which may have blurred differences in the quality of the internal representation (Just & Carpenter, 1985; Peruch & Savoyant, 1991).

Multisensory benefits on navigation performance

The second objective was to investigate whether audiovisual landmarks also improved navigation performance. Navigation performance was measured with a wayfinding task (i.e. task 4). Employing audiovisual landmarks instead of auditory or visual landmarks helped participants to select shorter routes and travel faster to the target landmarks. A very strong correlation between the length of the travelled route and the travel time indicates that participants travelled faster because they selected shorter routes. These results confirm the hypothesis that audiovisual landmark presentation improves navigation performance in virtual environments.

The literature on navigation state that information from our spatial memory enables us to find our way in our environment (Siegel & White, 1975). This statement is supported by strong correlations between draw map and navigation performance in real and virtual environments (Rovine & Weisman, 1989; Murakoshi & Kawai, 2000). These findings suggest that multisensory landmarks allowed our participants to navigate better because they improved the information about the environment in their spatial memory.

An alternative explanation, however, is that the presentation of audiovisual landmarks during the wayfinding task also helped participants to retrieve spatial information. In chapter 5, we showed that multisensory presentation during retrieval (i.e. during the wayfinding task) can be as effective as multisensory presentation during encoding (i.e. during the exploration of the maze). Further research will be required to disentangle these two explanations.

The role of distracter tasks and auditory landmarks in navigation

In this paragraph we would like to make a brief comment for putting our results in perspective. In the experiment described in this chapter (and also in the experiments described in chapters 5 and 6), participants had to perform a distracter task when they explored the environment and the items they had to remember. This distracter task was primarily used because it was shown that multisensory memory effects were larger when such a task was employed (Thompson & Paivio, 1994). The benefits of multisensory presentation may be lower in virtual environments which do not feature such a distracter task. However, while typical virtual environment applications do no feature such a task, they do distract navigators in other ways. That is, memorizing the environment is often not the user's main task in virtual environment applications. We think that this may play a similar role as the distracter task that was employed in our experiments.

What we think that also requires a brief comment are the results for spatial memory and navigation performance following exploration with auditory landmarks. In the introduction we argued that landmarks in the environment were not limited to the visual modality. We find no significant differences in spatial memory and navigation performance following exploration with auditory and visual landmarks. This suggests that non-visual landmarks can be as effective as visual landmarks for navigation.

Conclusion

In conclusion, our results establish that the presentation of audiovisual landmarks improves both spatial memory and navigation performance in comparison with the presentation of either auditory or visual landmarks. These results expand the benefits of multisensory presentation reported earlier in perception and memory.

Chapter 8. Discussion and conclusions

8.1 Summary and discussion of the findings

In the first chapter, we introduced the Trisensory Orientation Model (TOM) to discuss the (potential) benefits for the presentation of congruent visual, auditory and tactile information on navigation. The model assumes a sequence of processes at the perceptual, memory and navigation level. Perceptual task performance can benefit from multisensory presentation due to sensory interaction processes. Memory performance may benefit from better perceptual performance and from multisensory effects at the encoding and retrieval level. Because navigation performance relies on memory performance (identity and location), the result of this sequence is a better navigation performance.

With this model in mind, we formulated seven research questions at the perception, memory and navigation level. The research results described in chapters 2 to 7 show multisensory benefits at all levels. In this chapter, we discuss these findings, their mutual relations and their implications for TOM. It is organized as follows. First, we discuss the results of the experiments in relation to our research questions and hypotheses. Second, we discuss the model implications of these findings. Third, we will recommend future research.

Question 1: Are the effects of visuotactile presentation reflected in early EEG patterns?

It is of interest to know if electroencephalography (EEG) measurements show the effects of multisensory presentation on perceptual task performance. More particularly, early changes of EEG patterns are indicative of early low-level sensory integration processes in contrast with slower high-level cognitive processes. An early low-level sensory integration process should be less task-dependent and less susceptible to attentional processes (see Werkhoven, Van Erp & Philippi, 2010). Shams, Kamitani, Thompson & Shimojo (2001) reported the first evidence for early sensory integration processes by studying audiovisual stimuli. We wanted to know if these findings can be generalized to visuotactile stimuli.

We hypothesized that visuotactile interactions modulate activity along the visual cortex in a similar vein as audiovisual interactions, although the modulations may occur slightly later (about 10 ms) because tactile signals take longer than visual or auditory signals to reach the brain. In chapter 2, we employed (EEG) to examine the neural mechanisms underlying the touch-induced flash illusion. We found evidence for sensory interactions as early as 40-70 ms post stimulus offset. A subsequent comparison with EEGs of the sound-induced flash illusion which were reported earlier (Shams *et al.* 2001; Mishra, Martinez, Sejnowski & Hillyard, 2007) supports our hypothesis. Together, these results show that the effects of multisensory presentation are likely to be rooted in early sensory interaction (sensory integration) processes. We expect that early sensory interactions can play a role in a wide range of tasks, including navigation in virtual environments.

Question 2: Does congruent multisensory presentation improve perceptual task performance?

This question was investigated because we expect that spatiotemporal properties of objects and events in virtual (and real) environments are perceived better when presented in multiple sensory modalities, in particular when the sensory presentations are congruent.

Thusfar, multisensory perception has often been studied by presenting *incongruent* information to different sensory modalities (e.g. Ernst & Banks, 2002; Shams & Kamitani & Shimojo, 2000; 2002). The results of those studies show that *incongruent* presentation affects the mean (e.g. the perceived percept lies in between the information presented in each modality) and may improve the reliability (e.g. the variance in the multisensory responses is lower) of a multisensory percept. We complemented these results with experiments (see chapters 3 and 4) in which we investigated the effects of *congruent* bi- and tri-sensory presentation in temporal numerosity judgment.

We hypothesized that congruent multisensory presentation improved the observer's ability to estimate numerosity. In chapter 3 we found that congruent presentation of audition, vision and/or touch either accuracy (i.e., the average underestimation was up to 20% lower) or increased reliability (i.e., up to 30% lower variance) lower variance. In chapter 4, we found that audiovisual presentation increased accuracy by up to 20%, but we found no effect on the reliability. These findings largely confirm our hypothesis that congruent multisensory presentation yields better perceptual task performance.

Two observations are important for putting these results in perspective. First, the effects of congruent multisensory presentation on the mean can be considered advantageous, whereas classifying the mean response of incongruent multisensory presentation is controversial (is a percept that lies in between the information presented in each modality more or less 'correct' than a percept that equals that what was presented in a single modality?). Thus, multisensory benefits in numerosity estimation are evident in particular for congruent stimuli.

Second, in the experiments conducted in chapters 3 and 4 the benefits of congruent multisensory presentation were inversely related with unisensory performance. This means that when unisensory perceptual performance is (relatively) good the benefits of congruent multisensory presentation are (relatively) small. Based upon this observation, we expect that the benefits of the presentation of congruent spatiotemporal properties in virtual environments are only evident when the unisensory components are difficult to perceive.

Questions about multisensory memory

Now we have clearly shown the benefits of multisensory presentation at the perceptual task level, we want to show the subsequent effects on memory performance. Currently known multisensory memory benefits are limited to benefits of *audiovisual* semantic object presentation on the *encoding* of object *identity* memory. We want to know the effects of combinations with *tactile* presentation on the encoding as well the *retrieval* of object identity and object *location*. This is of interest because later navigation processes rely strongly on the recognition of identity and (relative) positions of objects in the environment.

Question 3: Do multisensory memory effects extend to visuotactile presentation?
Based upon the similar brain-imaging and behavioral results reported for audiovisual and visuotactile presentation we expected that multisensory memory effects also exist for visuotactile presentation. In chapter 5, where we presented visual, tactile and visuotactile Morse codes from different locations. The results show that visuotactile memory performance was about 10% higher than visual performance but not higher than tactile performance. This only partially confirms our hypothesis. We would like to stress, though, that multisensory performance increases relative to visual performance. Thus, from the perspective of 'visual' virtual environments this finding still has considerable impact.
The significant, but relatively limited multisensory memory effects observed in chapter 5 may have been caused by the fact that the participants did not know the meaning of the presented Morse codes. Indeed, the results of chapter 6 show that multisensory memory benefits for non-semantic Morse codes are smaller than for semantic items such as sounds and pictures of animals, musical instruments and other familiar objects (see Q6). Therefore, we recommend further investigation of visuotactile memory effects employing semantic instead of non-semantic items.

Question 4: Does multisensory presentation improve *retrieval* as well as encoding?
During the navigation of an environment our memory of that environment is accessed by encoding as well as retrieval processes. Encoding and retrieval processes are similar in the way that they both deal with (multisensory) object identity and object location information, and that they both match perceived information with that in memory. Therefore, we hypothesized beneficial multisensory effects on retrieval effects in addition to those on encoding.
In chapter 5, we found a 10% higher memory performance when visuotactile instead of visual Morse codes were presented during retrieval, when visual Morse codes were presented during encoding. This finding shows the existence of multisensory retrieval effects, confirming our hypothesis. In addition, we also found that the size of the reported multisensory retrieval effect was similar to that of the multisensory encoding effect. This result further supports the claim that encoding and retrieval are similar processes and the decision to represent encoding and retrieval as a single process in TOM (see Figures 1.3 and 1.5).

Question 5: Does multisensory presentation improve object *location* memory as well as object identity?
This question was investigated in chapter 5 (object location memory) and in chapter 7 (spatial memory about virtual environments). In chapter 5, visuotactile presentation increased object location memory performance by 10 to 20% in comparison with visual presentation. These results confirm the hypothesized existence of multisensory memory benefits on object location memory. Interestingly, multisensory memory benefits for object location memory were similar to and occurred in the same conditions as multisensory identity memory benefits.

In chapter 7, we investigated spatial memory by means of a draw map task following the exploration of an environment containing either audiovisual, auditory or visual objects. In that task, we found that multisensory presentation increases memory for (relative) object locations by approximately 75%. These benefits were much larger than the multisensory memory benefits for the number of object identities reported in the same task (which was only about 25%). We suggest that the multisensory memory benefit for the number of object identities was smaller because it was limited by a ceiling (i.e., the number of identities reported was close to the number of presented objects). Nonetheless, our results show that the size of the multisensory memory benefits on object identity and on object location may not need to be the same.

Question 6: Do multisensory memory benefits differ for congruent *semantic* and *non-semantic* items?

In chapter 5 we found significant, but relatively small, multisensory memory benefits for congruent but non-semantic items. We were interested to see if multisensory memory for *semantic* objects (more natural and familiar objects) would further improve. In chapter 6, we found that participants recalled 30 to 50% more *semantic* items when these items were presented both as sounds and pictures instead of either sounds or pictures. In contrast, no significant increase in the number of recalled items was found when non-semantic Morse codes were presented audiovisually instead of either only auditorily or only visually. The difference in multisensory memory effects on Morse codes reported in chapters 5 and 6 may be the result of differences in experimental design (such as a recognition versus a recall task). Further, it should be stressed that in chapter 5, 1080 trials were conducted with Morse codes while only 100 trials were conducted with Morse codes in chapter 6. Therefore, the multisensory memory effect reported in chapter 5, may have been too small to detect in chapter 6.

These results obtained in chapter 6 show that multisensory congruency by itself is not sufficient for multisensory benefits on memory; the congruent components need to carry actual semantic information. The results highlight the importance of semantic congruency over non-sematic (spatiotemporal) congruency for multisensory memory benefits.

Question 7: Do multisensory landmarks increase navigation performance?

Based on the observed effects of multisensory presentation on perceptual and memory tasks we hypothesized that the exploration of an environment containing audiovisual landmarks increases navigation performance compared with visual or auditory landmarks. In chapter 7, we show that participants indeed found their way faster (by approximately 20%) and took shorter routes to reach target locations after they had explored environments with audiovisual landmarks, which confirms our hypothesis A strong correlation between navigation time and route length indicates that participants found their way faster because they selected shorter routes.

The literature on spatial memory and navigation shows that spatial memory is a good predictor for navigation performance (Rovine & Weismann, 1989; Murakoshi & Kawai, 2000). Therefore, participants may have selected shorter routes because they had encoded more information about the environment during exploration with audiovisual instead of visual or auditory landmarks. Alternatively, it is also possible that the multisensory landmarks, which were also presented when participants had to find their way to the target locations, helped participants to retrieve the information that was stored during exploration.

8.2 Model implications of our experimental results

In this section we discuss the implications of our findings on the framework we proposed in the introduction. The experiments described in the individual chapters of this thesis show that multisensory presentation benefits perception, memory and navigation. In comparison with the presentation of information in the visual modality, presenting additional congruent information in the auditory and the tactile modality can improve the accuracy and/or reliability at the perceptual level, improve encoding and retrieval at the memory level, and increase navigation performance. These results are largely consistent with TOM. Therefore, we conclude that TOM is a useful framework.

Two observations are important for putting this framework into perspective. Firstly, multisensory interactions effects at the perceptual level (i.e., spatiotemporal congruency effects) only benefit navigation performance if the perceptual quality of each of the sensory modalities has not yet ‘saturated’. In chapters 3 and 4 we found that congruent multisensory presentation of non-semantic pulse series improves accuracy and reliability in perceptual tasks. However, in the chapters dealing with memory, we found only limited or even no effects of congruent multisensory presentation of non-semantic Morse codes. This may have been due to the fact that the Morse codes presented in the memory experiments were perfectly perceivable within each of the sensory modalities. Multisensory interaction at the perceptual level will only affect subsequent memory and navigation performance if the object perceived in the individual modalities is not yet perceived perfectly.

Secondly, in the memory and navigation experiments described in this thesis we presented both identities and locations in multiple sensory modalities. Our experiments show that the size of multisensory memory benefits for object identity and object location memory can differ. This may indicate that the encoding and retrieval of object identity information and object location information are independent processes and supports our decision to model *object identity memory* and *object location memor*y as separate components of *memory* in TOM. As we mentioned in chapter 5, it would be interesting to separately study the effects of multisensory object identity information on object identity memory and the effects of multisensory object location information on object location memory. We expect that these effects may be seperable.

8.3 Concluding remarks

The overarching purpose of this thesis was to investigate the beneficial effects of congruent visual, auditory, and tactile presentation on navigation performance. Overall, the results show that multisensory presentation is advantageous. An important aspect of this finding is that multisensory performance was always equal to or higher than the best unisensory performance. Therefore, we expect that presenting multisensory information in virtual environment applications has only positive effects. Of course, the cost of implementing multisensory environments (e.g. investment) should be lower than the benefits (e.g. reduction of training time; return on investment).

While we report positive effects of multisensory presentation in all chapters of this thesis, we have only conducted a single experiment which shows that multisensory presentation actually improves navigation. It is yet unclear to what extent the effects of this experiment will carry over to a wider range of virtual environmental applications. Further research is required to provide adequate guidelines when and when not the implementation of (congruent) multisensory presentation is advisable. We have two comments relevant to this issue.

Firstly, the investigation of multisensory presentation in virtual environments in this thesis was limited to local landmarks (i.e. landmarks that are only observable from their immediate surroundings). Many more elements in a virtual environment can be made multisensory and the benefits thereof need not necessarily be the same to those for multisensory local landmarks. Global landmarks (i.e., landmarks that are observable from any location within a large area; e.g. a church tower), environments (e.g. urban or rural), and displacement (e.g. footsteps and/or engine sounds) can all easily be presented congruently in audition and vision. Displacement could also be presented to touch. Even if memory and navigation benefits for these elements are not as large as those we report for local landmarks, the benefits may all add up when everything is presented congruently in multiple sensory modalities.

A second and final consideration for the implementation of congruent multisensory presentation is that sensory channels only have a limited bandwidth. This means that if, for example, one chooses to use auditory and/or tactile modalities to present information complementary and congruent with the visual modality - and the auditory and tactile resources are used to the full extent - that the auditory and tactile modalities cannot be used to convey other and perhaps more beneficial information. For instance, the presentation of auditory information may interfere with our ability to communicate. On the other hand, multisensory presentation gives the navigator the option to ad hoc use a sensory modality for other purposes without fully reducing navigation performance.

In conclusion, our experiments show that multisensory presentation benefits perceptual, memory, and navigation performance. In addition, our experiments have provided further insight in the mechanisms underlying multisensory interactions. They imply that the presentation of multisensory information in virtual environments may help to overcome some of the current difficulties associated with navigating virtual environments. We expect that multisensory presentation increases a navigator's ability to learn information about the environment and helps him or her to select better routes, thus improving the navigator's ability to find his way in a virtual environment.

List of publications and presentations

Journal papers

Philippi, T.G., Van Erp, J.B.F. and Werkhoven, P.J. (2008). Multisensory Temporal Numerosity Judgment. *Brain Research, 1242*, 116-125.

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. (*in press*). Multisensory effects differ for counting small and large pulse numbers. *Seeing and Perceiving*

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. (*under revision*). Visuotactile memory of object identity and object location.

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. (*submitted*). Experience helps to distinguish flash illusions.

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. (*in preparation*). Multisensory memory effects differ for familiar objects and Morse codes.

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. (*in preparation*). Audiovisual navigation in virtual mazes.

Werkhoven, P.J., Van Erp, J.B.F. and Philippi, T.G. (2009). Counting visual and tactile events: the effect of attention on multisensory integration. *Attention, Perception & Psychophysics, 71(8)*, 1854-1861

Conference proceedings

Philippi, T.G., Van Erp, J.B.F. and Werkhoven, P.J. (2010). Is the touch-induced illusory flash distinguishable from a real flash? *LNCS* 6192, 418-425. Heidelberg: Springer.

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. (*in preparation*). Early sensory interactions of the touch-induced flash illusion: An EEG study.

Refereed abstracts

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P, Bronkhorst, A. Does multisensory integration improve temporal numerosity judgments? *Poster presentation, International Multisensory Research Forum, Sydney, 5-7 July, 2007*

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. Is bias and variance of multimodal temporal numerosity judgement consistent with Maximum Likelihood Estimation? *Poster presentation, International Multisensory Research Forum, Hamburg, 15-19 July, 2008*

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. Touch alters activity along the human visual cortex. *Oral presentation, Dag van de Perceptie, Soesterberg, 6 November, 2008*

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. Benefits of multisensory memory encoding and retrieval *Poster presentation, NVP Winter Conference on Cognition, Brain, and Behaviour, December 18-19, 2009*

Philippi, T.G., Van Erp, J.B.F., Werkhoven, P. Audiovisual navigation in virtual mazes. *Oral presentation, Dag van de Perceptie, Eindhoven, 8 April, 2011*

References

Aguirre, G.K. D'Esposito, M. (1997). Environmental knowledge is subserved by seperable dorsal/ventral neural areas. *The Journal of Neuroscience 17(7),* 2512-2518.

Alais, D., Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257-262.

Allan, L.G. (1979). The perception of time. *Perception & Psychophysics, 26(5),* 340-354

Andersen, T.S., Tiippana, K. and Sams, M. (2004). Factors influencing audiovisual fusion and fission illusions. *Cognitive Brain Research, 21(3)*, 301-208.

Andersen, T.S., Tiippana, K., Sams, M. (2005). Maximum Likelihood Integration of rapid flashes and beeps. *Neuroscience Letters, 380*, 155-160.

Arden, G.B., Wolfe, J.E., Messiter, C. (2003). Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. *Vision Research*, *43,* 2469-78

Ardito, C., Costabile, M., De Angeli, A., Pittarello, F. (2007). Navigation help in 3D worlds: some empirical evidences on use of sound. *Multimedia Tools and Applications, 33(2),* 201-216.

Azuma, R. (1997) A survey of augmented reality. *Presence: Teleoperators & Virtual Environments 6 (4):* 355-386

Baddeley, A.D. (1986). *Working Memory*. Oxford: Oxford University Press.

Bakker, N.H. (2001). *Spatial orientation in Virtual Environments.* Doctoral thesis, Delft Univeristy of Technology. Delft, The Netherlands: DUP Science.

Bauer, M., Oostenveld, R., Fries, P. (2009). Tactile stimulation accelerates behavioral responses to visual stimuli through enhancement of occipital gamma-band activity. *Vision Research, 39*, 931-942.

Boyd, J.R. (1987). *A discourse on winning and losing* [Briefing slides]. Maxwell Air Force Base, AL: Air University Library. (Document No. M-U 43947)

Behar, I., Bevan, W. (1961). The perceived duration of auditory and visual intervals: Cross-modal comparison and interaction. *American Journal of Psychology, 74*, 17-26.

Benevento, L., Fallon, J., Davis, B.J., Rezak, M. (1977). Auditory visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Experimental Neurology, 157*, 849-872

Bernstein, I.H., Clark, M.H., Edelstein, B.A. (1969). Effects of an auditory signal on visual reaction time. *Journal of Experimental Psychology*, *80*, 567-569.

Bhattacharya, J., Shams, L., Shimojo, S. (2002). Sound-induced illusory flash perception: role of gamma band responses. *Neuroreport, 13(4),* 1727-1730

Boring, E.G. (1942). *Sensation and Perception in the History of Experimental Psychology*. D. Appleton-Century Company, New York, London.

Bowman, D.K. (1995). International survey: virtual-environment research. *IEEE Computer, June,* 57-65.

Bresciani, J.-P., Dammeier, F., Ernst, M.O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*, 554-664.

Bresciani, J.-P. and Ernst, M.O. (2007). Signal reliability modulates auditory-tactile integration for event counting. *Neuroreport, 18(11)*, 1157-1161.

Bresciani, J. P., Ernst, M. O., Drewing, K., Bouyer, G., Maury V., Kheddar, A. (2005). Feeling what you hear: auditory signals can modulate tactile tap perception. *Experimental Brain Research, 162,* 172-180.

Burigat, S., Chittaro, L. (2007). Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids. *International Journal of Human Computer Studies, 65(11),* 945-958.

Burr, D., Banks, M., Morrone, M.C. (2009) Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research, 198*, 49-57.

Caduff, D., Timpf, S. (2006). A framework for assessing the salience of landmarks for wayfinding tasks. *Cognitive Processes, 7(S1),* S23.

Cheatham, P.G., White, C.T. (1952). Temporal numerosity I: Perceived number as a function of flash number and rate. *Journal of Experimental Psychology, 44(6)*, 283–287.

Cheatham, P.G., White, C.T. (1954). Temporal numerosity III: Auditory perception of number. *Journal of Experimental Psychology, 47(6)*, 425–428.

Chen, K.-M., Yeh, S.-L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychololgica, 130*, 225-234.

Courtney, J.R., Motes, M.A., Hubbard, T.L. (2007). Multi- and unisensory visual flash illusions. *Perception*, *36*, 516–524.

Darken, R.P., Sibert, J.L. (1996). Navigating in large virtual worlds. *The International Journal of Human-Computer Interaction, 8(1),* 49-72.

De Lange, H. Dzn. (1952). Experiments on Flicker and Some Calculations on an Electrical Analogue of the Foveal Systems. *Physica*, *18*, 935-950.

Delogu, F., Raffone, A., Olivetti Belardinelli, M. (2009). Semantic encoding in working memory: Is there a (multi)modality effect? *Memory, 17(7),* 655-663.

Dewhurst, S.A., Knott, L.M. (2010). Investigating the encoding-retrieval match in recognition memory: Effects of experimental design, specificity, and retention interval. *Memory & Cognition, 38,* 1101-1109.

Doehrmann, O., Naumer, M.J. (2008). Semantics and the multisensory brain: How meaning modulates processes of audio-visual integration. *Brain Research, 1242,* 136-150

Driver, J., Spence, C. (1998). Crossmodal Attention. *Current Opinion in Neurobiology, 8(2),* 245-253.

Droit-Volet, S., Meck, W.H., Penny, T.B. (2007). Sensory perception and time perception in children and adults. *Behavorial Processes*, *74(2)*, 244-250.

Efron, R. (1970a). The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologica, 8*, 37-55.

Efron, R. (1970b). The minimum duration of a perception. *Neuropsychologica, 8*, 57-63.

Efron, R. (1973). An invariant characteristic of perceptual systems in the time domain. In *Attention and Performance IV*. Lornblum, S., ed. Academic Press, New York.

Eichenbaum, H., Cohen, N.J. (2001). *From Conditioning to Conscious Recollection: Memory Systems of the Brain.* New York: Oxford University Press.

Ellis, S.R. (Ed.) (1993). *Pictorial communication in virtual and real environments.* Taylor & Francis Ltd., London.

Ernst, M.O. (2006). A Bayesian view on multimodal cue integration. In *Human Body Perception From The Inside Out*. Knoblich G., Thornton, I.M., Grosjean, M., Shiffrar, M., eds. New York, Oxford University Press.

Ernst, M.O., Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

Ernst, M.O., Bülthoff, H.H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Science, 8,* 162-169.

Ernst, M.O., Lange, C., Newell, F.N. (2007). Multisensory recognition of actively explored objects. *Canadian Journal of Experimental Psychology, 61(3),* 242-253

Falchier, A., Clavagnier, S., Barone, P., Kennedy, H. (2002). Anatomical evidence of multimodal intergration in primate striate cortex. *Journal of Neuroscience, 22(13),* 5749-5759

Fries, W. (1984). Cortical projections to the superior colliculus in the macaque monkey: a retrograde study using horseradish peroxidase. *Journal of Computational Neurology, 230,* 55-67

Foxe, J.J., Simpson, G.V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing. *Experimental Brain Research, 142,* 139-150

Fu, K.-M.G., Johnston, T.A., Shah, A.S., Arnold, L., Smiley, J., Hackett, T.A., Garraghty, P.E., Schroeder, C.E. (2003). Auditory cortical neurons respond to somatosensory stimulation. *Journal of Neuroscience, 23(20),* 7510-7515

Gallace, A., Tan, H.Z., Spence, C. (2007). Multisensory numerosity judgments for visual and tactile stimuli. *Perception & Psychophysics*, *69(4)*, 487-501.

Garner, W.R., Gottwald, R.L. (1968). The perception and learning of temporal patterns. *Quarterly Journal of Experimental Psychology, 20*, 97-107.

Gault, R.H., Goodfellow, L.D. (1938). An empirical comparison of audition, vision and touch in the discrimination of temporal patterns and ability to reproduce them. *The Journal of General Pscyhology*, *18*, 41-47.

Gepshtein, S., Banks, M.S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, *13*, 483-488.

Gescheider, G.A., Valetutti, A.A., Padula, M.C., Verrillo, R.T. (1992). Vibrotactile forward masking as a function of age. *Journal of the Acoustical Society of America*, *91(3)*, 1690-1696.

Ghazanfar, A.A., Schroeder, C.E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Science, 11,* 473-490.

Gibbon, J., Church, R. M. and Meck, W. H. (1984). Scalar timing in memory. *Annals of the New York Academy of Sciences, 423*, 52–77.

Gielen, S.C.A.M., Schmidt, R.A., van den Heuvel, P.J.M. (1983). On the nature of intersensory facilitation of reaction time. *Perception & Psychophysics, 34*, 161-168.

Giller, S., Mallot, H.A. (1998). Navigation and acquisition of spatial knowledge in a virtual maze. *Journal of Cognitive Neuroscience, 10(4),* 445-463.

Goldin, S.E., Thorndyke, P.W. (1982). Simulating navigation for spatial knowledge acquisition. *Human Factors, 24(4),* 457-471.

Goldstone, S. (1968). Variability of temporal judgment: intersensory comparisons and sex differences. *Perception and Motor Skills, 26(1),* 211.

Goldstone, S., Boardman, W.K., Lhamon, W.T. (1959). Intersensory comparisons of temporal judgments. *Journal of Experimental Psychology, 57*, 243-248

Goldstone, S., Goldfarb, J. (1963). Judgment of filled and unfilled duration: Intersensory effects. *Perceptual and Motor Skills*, *17*, 763-774

Goldstone, S., Lhamon, W.T. (1972). Auditory – visual differences in human judgment. *Perceptual and Motor Skills*, *34*, 623-633

Green, D.M., Swets, J.A. (1966). *Signal Detection Theory and Psychophysics.* New York: Wiley. ISBN 0-471-32420-5.

Grondin, S. (2003). Sensory modalities and temporal processing. In: H Helfrich (Ed.), *Time and mind II* (pp. 61-77). Göttingen: Hogrefe & Huber.

Gunther, R., Kazman, R., MacGregor, C. (2004). Using 3D sound as a navigational aid in virtual environments. *Behaviour and Information Technology, 23(6),* 436-446.

Handel, S., Buffardi, L. (1969). Using several modalities to perceive one temporal pattern. *Quarterly Journal of Experimental Psychology, 21*, 256-266.

Hasher, L. Zacks, R.T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356-388.

Hershenson, M. (1962). Reaction time as a measure of intersensory facilitation. *Journal Experimental Psychology, 63*, 289-293.

Hötting, K., Friedrich, C.K., Röder, B. (2009). Neural correlates of cross-modally induced changes in tactile awareness. *Journal of Cognitive Science, 21(12),* 2445-2461

Hötting, K. and Röder, B. (2004). Hearing cheats touch, but less in congenitally blind than in sighted individuals. *Psychoogical Science, 15*, 60-64.

Ishiwaka, T., Montello, R.D. (2006). Spatial knowledge acquisition from direct experience in the environment: individual differences in the development of metric knowledge and the integration of separately learned places. *Cognitive Psychology, 52,* 93-129.

Janzen, G. (2006). Memory for object location and route direction in virtual large-scale space. *The Quarterly Journal of Experimental Psychology, 59(3),* 493-508

Just, M.A., Carpenter, P.A. (1985). Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92,* 137-172.

Kanabus, M., Szelag, E., Rojek, E., Poppel, E. (2002). Temporal order in different sense modalities. *Journal of Experimental Psychology, 62*, 423-432.

Kaufman, E., Lord, M., Reese, T. and Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498-525.

Kaur, K. Sutcliffe, A.G., Maiden, N.A.M. (1999). Towards a better understanding of usability problems with virtual environments. *Proceedings of the IFIP TC.13 International Confgerence on Human-Computer Interaction (INTERACT'99), Edinburgh, UK., 527-535.*

Keller, F.S., Schoenfeld, W.N. (1944). Studies in international morse code: III. The efficiency of the code as related to errors made during learning. *Journal of Applied Psychology, 28*, 254–266.

Keller, F.S., Taubman, R.E. (1943). Studies in international morse code: II. errors made in code reception. *Journal of Applied Psychology, 27*, 504–509.

Kent, C., Lamberts, K. (2008). The encoding–retrieval relationship: Retrieval as mental simulation. *Trends in Cognitive Sciences***,** *12,* 92-98.

Klink, P.C., Montijn, J.S., van Wezel, R.J.A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attenention*, *Perception & Psychophysics, 73*, 219-236

Kolers, P.A., Brewster, J.M. (1985). Rhythms and responses. J. Exp. Psychol. Human 11, 150-167.

Kuipers, B. (1982). The "map in the head" metaphor. *Environment and Behaviour, 14(2),* 202-220.

Landy, M.S., Maloney, L.T., Johnston, E.B., Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *25*, 389-412.

Laurenti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H., Wallace, M.T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research, 158,* 405-414.

Lechelt, E.C. (1975). Temporal Numerosity Discrimination: Intermodal Comparisons Revisited. *British Journal of Psychology*, *66(1)*, 101-108.

Lehmann, S., Murray, M.M. (2005). The role of multisensory memories in unisensory object discrimination. *Cognitive Brain Research*, *24(2)*, 326-334.

Lehnert, G., Zimmer, H.D. (2006). Auditory and visual spatial working memory. *Memory & Cognition, 34(5),* 1080-1090.

Lehnert, G., Zimmer, H.D. (2008a). Modality and domain specific components in auditory and visual working memory tasks. *Cognitive processes, 9,* 53-61.

Lehnert, G., Zimmer, H.D. (2008b). Common coding of auditory and visual spatial information in working memory. *Brain Research, 1230,* 158-167.

Levinson, J.Z. (1968). Fusion Flicker Phenomena. *Science, 160(3823)*, 21-28.

Lhamon, W.T., Goldstone, S. (1974). Studies of auditory – visual differences in human time judgment 2. More transmitted information with sounds than lights. *Perceptual and Motor Skills*, *39*, 295-307.

Logie, R. H. (1995). *Visuo-spatial Working Memory.* Hove, UK: Lawrence Eribaum Associates.

Macaluso, E. (2006). Multisensory processing in sensory-specific cortical areas, *The Neuroscientist, 12,* 327-338.

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264(5588),* 746-748.

Mishkin, M., Ungerleider, L.G., Macko, K.A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Cognitive Neuroscience 6,* 414-417.

Mishra, J, Martinez, A., Hillyard, S.A. (2008). Cortical processes underlying sound-induced flash fusion. *Brain Research, 1242,* 102-115.

Mishra, J., Martinez, A., Sejnowski, T.J., Hillyard, S.A. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *Journal of Neuroscience*, *27(15)*, 4120-4131.

Montello, R.D. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments. In: J.M. Egenhofer and G.R. Golledge, Editors, *Spatial and temporal reasoning in geographinc information systems*, Oxford University Press, New York, pp. 143–154.

Morell, L.K. (1968). Cross-modality effects upon choice reaction time. *Psychonomic Science*, *11*, 129-130.

Murakoshi, S., Kawai, M. (2000). Use of knowledge and heuristics for wayfingind in an artificial environment. *Environment and Behaviour, 32(6),* 756-774.

Murray, M.M., Michel, C.M., Grave de Peralta, R., Ortigue, S., Brunet, S., Gonzales Andino, S., Schnider, A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. *Neuroimage, 21(1),* 125-135.

Moscovitch, C., Kapur, S., Kohler, S., Houle, S. (1995). Distinct neural correlates of visual long-term memory for spatial location and object identity: a positron emission tomography study in humans. *Proceedings of the National Academy of Sciences USA 92*, 3721-3725.

Nickerson, R.S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review*, *80*, 489-509.

Noesselt, T., Rieger, J.W., Schoenfeld, M.A., Kanowski, M., Hinrichs, H., Heinze, H.-J., Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary cortices. *Journal of Neuroscience, 27(42)*, 11431-11441.

Osinga, F. (2006). *Science Strategy and War, The Strategic Theory of John Boyd.* Abingdon, UK: Routledge.

Paivio, A (1971). *Imagery and verbal processes.* New York: Holt, Rinehart, and Winston.

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

Palij, M., Levine, M., Kahan, T. (1984). The orientation of cognitive maps. *Bulletin of the Psychonomic Society, 22,* 105-108.

Péruch, P. and Savoyant, A. (1991). Conflicting spatial frames of reference in a locating task. In: Logie, R.H. and Denis, M., Editors, 1991. *Mental images in human cognition*, Elsevier, Amsterdam, pp. 47–55.

Philippi, T.G., Van Erp, J.B.F. and Werkhoven, P.J. (2008). Multisensory Temporal Numerosity Judgment. *Brain Research, 1242*, 116-125.

Philippi, T.G., Van Erp, J.B.F. and Werkhoven, P.J. (2010). Is the touch-induced illusory flash distinguishable from a real flash? *LNCS* 6192, 418-425. Heidelberg: Springer.

Postma, A., Kessels, R.P.C., Van Asselen, M. (2004). The neuropsychology of object location memory. In: C. Auen (Ed.), *Remembering where: Advances in understanding spatial memory.* Lawrence Erlbaum associateds: Mahwah, NJ.

Postma, A., Kessels, R.P.C., Van Asselen, M. (2008). How the brain remembers and forgets where things are: The neurocognition of object-location memory. *Neuroscience and Biobehavioral Reviews, 32,* 1339-1345.

Presson, C.C., Montello, D.R. (1988). Points of reference in spatial cognition: Stalking the elusive landmark. *British Journal of Developmental Psychology, 6,* 371-381.

Presson, C.C., DeLange, N., Hazelrigg, M.D. (1989). Orientation specificity in spatial memory: what makes a path different from a map of the path? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(5),* 887-897.

Rockland, K.S., and Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology, 50,* 19-26.

Rovine, M.J., and Weisman, G.D. (1989). Sketch-map variables as predictors of way-finding performance. *Journal of Environmental Psychology, 9,* 217-232

Riecke, B.E., Van Veen, H.A.H.C., Bülthoff, H.H. (2002). Visual homing is possible without landmarks: A path integration study in virtual reality. *Presence, 11(5),* 443-473.

Richardson, A.E., Montello, D.R., Hegarty, M. (1999). Spatial knowledge acquisition from maps and from navigation in real and virtual environments. *Memory & Cognition, 27(4),* 741-750.

Ruddle, R.A., Lessels, S. (2006). For efficient navigational search, humans require full physical movement, but not a rich visual scene. *Pscyhologial Science* 17: 460-465

Sathian, K., Zangaladze, M.D. (2001). Feeling with the Mind's Eye: the Role of Visual Imagery in Tactile Perception. *Optometry and Vision Science, 78,* 276-281.

Sauvan, X.M., Peterhans, E. (1999). Orientation constancy in neurons of the monkey visual cortex. *Visual Cognition, 6(1),* 43-54

Schacter, D.L. (2007). Memory: Delineating the core. In: Roediger, H.L., Dudai, Y., Fitzpatrick, S.M. (Eds.), *Science of Memory: Concepts,* New York: Oxford University Press.

Schaefer, M., Flor, H., Heinze, H.-J., and Rotte, M. (2005). Dynamic modulation of the primary somatosensory cortex during seeing and feeling of a touched hand. *Neuroimage, 29,* 587-592

Schroeder, C.E., Lindsley, R.W., Specht, C., Marcovici, A., Smiley, J.F., Javitt, D.C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. *Journal of Neurophysiology, 85(3),* 1322-1327

Schroeder, C.E., Foxe, J.J., (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Cognitive Brain Research, 14,* 187-198

Schroeder, C.E., Foxe, J.J., (2005) Multisensory contributions to low-level 'unisensory' processing. *Current Opinions in Neurobiology*, *15,* 454-458

Seashore, H., Kurtz, A.K. (1943). Analysis of errors in copying code. *OSRD Report 4010*.

Seitz, A.R., Kim, R., Shams, L. (2006). Sound facilitates visual learning. Current Biology, 16(14), 1422-7.

Sekuler R., Sekuler A.B., Lau, R. (1997). Sound alters visual motion perception. *Nature, 385,* 308

Shams, L., Iwaki, S., Chawla, A., Bhattacharya, J., (2005). Early modulation of visual cortex by sound: an MEG study. *Neuroscience Letters, 378,* 76-81

Shams, L., Kamitani, Y., Shimojo, S. (2000). What you see is what you hear. *Nature, 408*, 788.

Shams, L., Kamitani, Y., Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*, 147-152.

Shams, L., Kamitani, Y., Thompson, S, Shimojo, S. (2001). Sound alters visual evoked potential in humans. *Neuroreport*, *12(17)*, 3849-3852.

Shams, L., Ma, W.J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport, 16(17)*, 1923-1927.

Shams, L., Seitz, A. (2008). Benefits of multisensory learning. *Trends in Cognitive Science, 12(11),* 411-417.

Shelton, A.L., McNamara, T.P. (2004). Orientation and perspective dependence in route and survey learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 158-170.

Shipley, T. (1964). Auditory-flutter driving of visual flicker. *Science*, *145(3638)*, 1328-1330.

Siegel, A.W., White, S.H. (1975). The development of spatial representations of large-scale environments. In: Reese, H.W., editor, *Advances in Child Development and Behavior*, volume 10, New York: Academic press, pp. 9-55.

Slotnick, S.D. (2004). Visual memory and visual perception recruit common neural substrates. *Behavioral and Cognitive Neuroscience Reviews, 3,* 207–221

Snodgrass, J. G., Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174-215.

Souman, J.L., Robuffo Giordano, P., Schaiger, M., Frissen, I., Thümmel, T., Ulbrich, H., De Luca, A., Bülthoff, H.H., Ernst, M.O. (2011). CyberWalk: Enabling unconstrained omnidirectional walking through virtual environments. *ACM Transactions on Applied Perception, 8(4),* 24.

Spence, C., Driver, J. (2004). *Crossmodal space and crossmodal attention.* Oxford: Oxford University Press.

Stein, B.E., London, N., Wilkinson, L.K., Price, D.D. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience, 8*, 497-506.

Stein, B.E., Meredith, M.A. (1993). *The merging of the senses*. Cambridge, MA, MIT Press.

Squire, L.R., (1987). *Memory and Brain.* New York: Oxford University Press.

Suid, C., Bonneel, N., Viaud-Delmon, I. (2009). Integration of auditory and visual information of realistic objects. *Experimental Brain Research, 194,* 91-102.

Taubman, R.E. (1950a). Studies in judged number: I. the judgment of auditory number. *Journal of General Psychology*, *43*, 195–219.

Taubman, R.E. (1950b). Studies in judged number: II. the judgment of visual number. *Journal of General Psychology*, *43*, 167–194.

Teder-Sälejärvi, W.A., McDonald, J.J., Di Russo, F., and Hillyard, S.A. (2002). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Cognitive Brain Research, 14,* 106-114.

Thompson, V.A. Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology, 48(3),* 380-398.

Thorndyke, P. (1981). Spatial cognition and reasoning. In: *Cognition, social behavior, and the environment* (Harvey J., ed.). Hillsdale, NJ: Erlbaum.

Tolman, C.E. (1948). Cognitive maps in rats and man. *Psychological Review, 55,* 189-208.

Treisman, M., (1963). Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psvchological Monographs, 77,* 13, Whole No. 576.

Tulving, E., Pearlstone, Z., (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381-391.

Tulving, E., Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

Ungerleider, L.G., Haxby, J.V. (1994). “What” and “where” in the human brain. *Current Opinions in Neurobiology* 4:157-165.

Van Erp, J.B.F. (2007). *Tactile displays for navigation and orientation: perception and behaviour.* Doctoral thesis, Utrecht University, Utrecht, The Netherlands.

Van Erp, J.B.F., Eriksson, L., Levin, B., Carlander, O., Veltman, J.A., & Vos, W.K. (2007). Tactile cueing effects on performance in simulated aerial combat with high acceleration. *Aviation Space and Environmental Medicine, 78(12),* 1128-1134.

Van Erp, J.B.F. and Spapé, M.M.A. (2008). Time-shrinking and the design of tactons. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS* 5024, 289-294. Heidelberg: Springer.

Van Erp, J.B.F. and Van Veen, H.A.H.C. (2006). Touch down: The effect of artificial touch cues on orientation in microgravity. *Neuroscience Letters, 404*, 78-82.

Van Erp, J.B.F., Verschoor, M.H., (2004). Cross-modal visual and vibrotactile tracking. *Applied Ergonomics, 35 (2),* 105-112.

Van Erp, J.B.F., Werkhoven, P.J. (2004). Vibro-tactile and visual asynchronies: Sensitivity and consistency. *Perception, 33*, 103-111.

Van Wassenhove, V., Buonomano, D.V., Shimojo S., Shams, L. (2008) Distortions of subjective time perception within and across senses. *PLoS ONE, 3(1)*, e1437.

Violentyev, A., Shimojo, S. and Shams, L. (2005). Touch-induced visual illusion. *Neuroreport*, *16(10)*, 1107-1110.

Von Kriegstein, K., Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS Biology, 4*, e326.

Vroomen, J., De Gelder, B. (2000). Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *Journal of Experimental Pscyhology: Human Perception and Performance*, *26(5)*, 1583–1590.

Walker, J.T., Scott, K.J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *Journal Experimental Psychology: Human Perception & Performance, 7*, 1327-1339.

Warren, W.H., Morris, M.W., Kalish,M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 646-660.

Watkins, S., Shams, L., Tanaka, S., Haynes, J.D., Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, *31*, 1247-1256.

Wearden, J.H., Edwards H., Fakhri, M., Percival, A. (1998). Why sounds are judged longer than lights: application of a model of the internal clock in humans. *Quarterly Journal of Experimental Psychology, 51*, 97-120.

Wieskrant, L., Neuroanatomy of memory and amnesia:"a case for multiple memory systems. *Human Neurobiology, 6,* 93-105.

Welch, R.B., Warren, D.H. (1980). Immediate perceptual response to inter-sensory discrepancy. *Psychological Bulletin, 88,* 638-667.

Wells, E.F., Bernstein, G.M., Scott, B.W., Bennet, P.J., Mendelson, J.R. (2001). Critical flicker frequency responses in the virtual cortex. *Experimental Brain Research*, *139*, 106-110.

Werkhoven, P.J., Van Erp, J.B.F. and Philippi, T.G. (2009). Counting visual and tactile events: the effect of attention on multisensory integration. *Attention, Perception & Psychophysics, 71(8)*, 1854-1861

White, C.T., Cheatham, P.G. (1959). Temporal numerosity IV: A comparison of the major senses. *Journal of Experimental Pscyhology, 58(6)*, 441–444.

Wickens, C.D. (1992). *Engineering psychology and human performance (2nd edition).* New York: Harper Collins.

Wickens, C.D., Hollands, J.G. (2000). *Engineering psychology and human performance (3rd edition).* Upper Saddle River, NJ: Prentice Hall

Wilson, P.N., Foreman, N., Tlauka, M. (1997). Transfer of spatial information from a virtual to a real environment. *Human Factors,* 39(4), 526-531.

Wolbers, T., Hegarty, M. (2010). What determines our navigational abilities? *Trends in Cognitive Sciences, 14(3),* 138-146.

Woods, A.T., Newell, F.N. (2004). Visual, haptic and cross-modal recognition of objects and scenes. *Journal of Physiology Paris, 98(1-3),* 147-159

Wu, W.C., Basdogan, C., Srinivasan, M.A. (1999). Visual, haptic and bimodal perception of size and stiffness in virtual environments. *Proceedings of the ASME Dynamic Systems and Control Division, 67*, 19-26.

Yuvall-Greenberg, S., Deouell, L. Y. (2007). What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. *Journal of Neuroscience, 27(5)*, 1090-1096.

Yuvall-Greenberg, S., Deouell, L.Y. (2009). The dog's meow: asymmetrical interaction in cross-modal recognition. *Experimental Brain Research (2009), 193,* 603-614.

Appendix

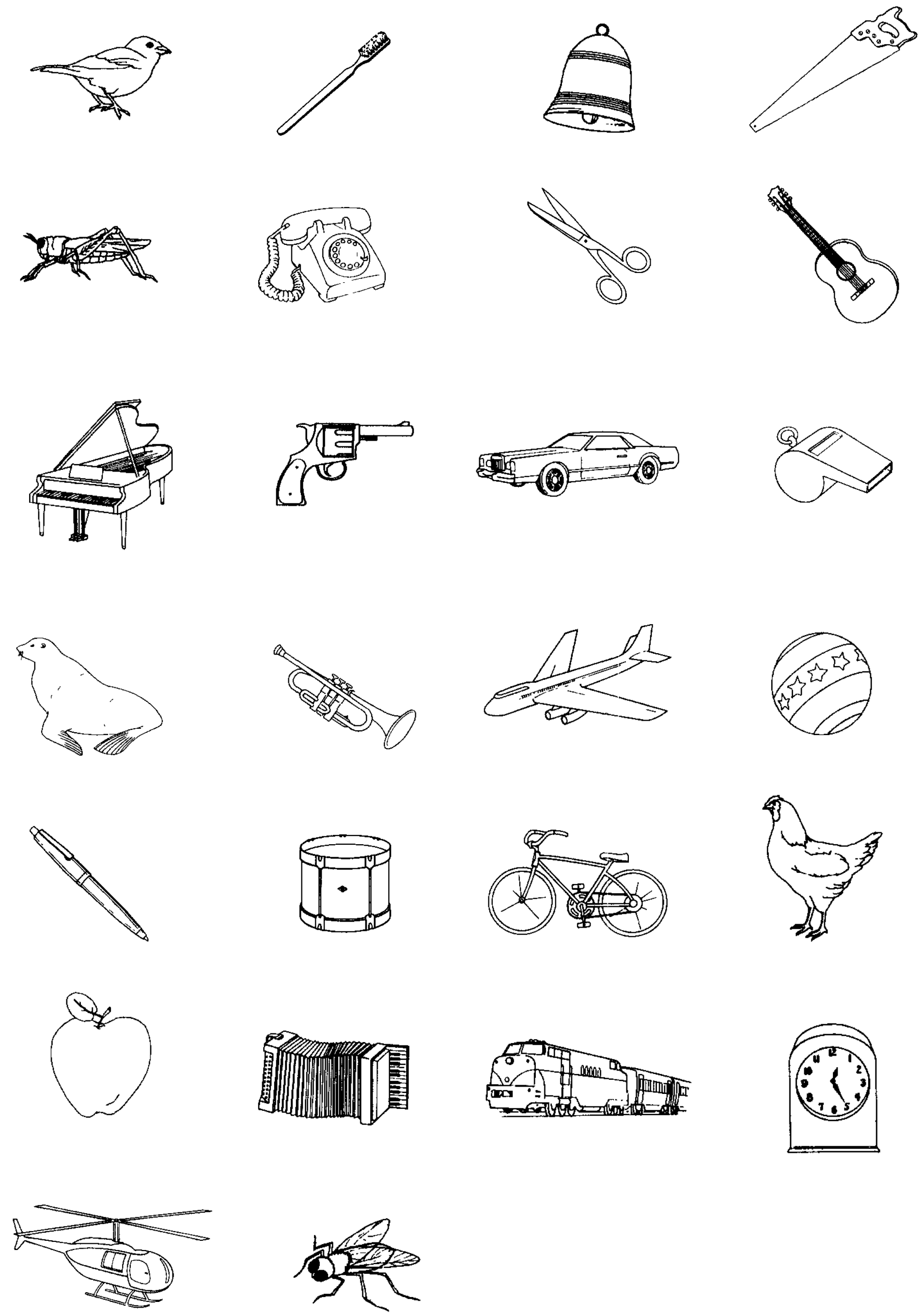
In chapter 6 and 7 stimulus sets consisting of semantic (i.e. familiar objects) and non-semantic objects (i.e. Morse codes) were presented. Table AI lists the items presented in chapter 6. This table is followed by a list of all images of the semantic items presented in that chapter. A second list displays all images of the landmarks as presented in chapter 7.

Table AI. Semantic and non-semantic items presented in the experiments of chapter 6

Semantic items (pictures)	Semantic items (sounds)	Non-semantic items
Bird	Chirping bird	— — —
Toothbrush	Teeth brushing	• — — •
Bell	Resonating bell	• —
Handsaw	Sawing	• — —
Grasshopper	Stridulation	• — •
Telephone	Telephone ringing	— • — —
Scissors	Scissors opening and closing	• • — —
Guitar	Striking of acoustic guitar strings	• • • —
Car	Car motor sound moving away	— •
Piano	Piano song	• • •
Gun	Multiple gun shots	— — • •
Whistle	Whistling	• • — •
Sealion	Barking sealion	— • • —
Trumpet	Trumpet tune	— — — •
Airplane	Jet-airplane landing	•
Ball	Ball hitting the ground multiple times	— —
Ballpoint pencil	Ballpoint clicking and writing	• — • •
Drum	Strum	• •
Bicycle	Bicycle bell	• — • —
Chicken hen	Chuckling hen	— • —
Accordion	Accordion tune	
Apple	Person biting an apple	
Train	Electric diesel train sound	
Clock	Ticking clock	
Helicopter	Helicopter rotor sound	
Fly	Buzzing fly	

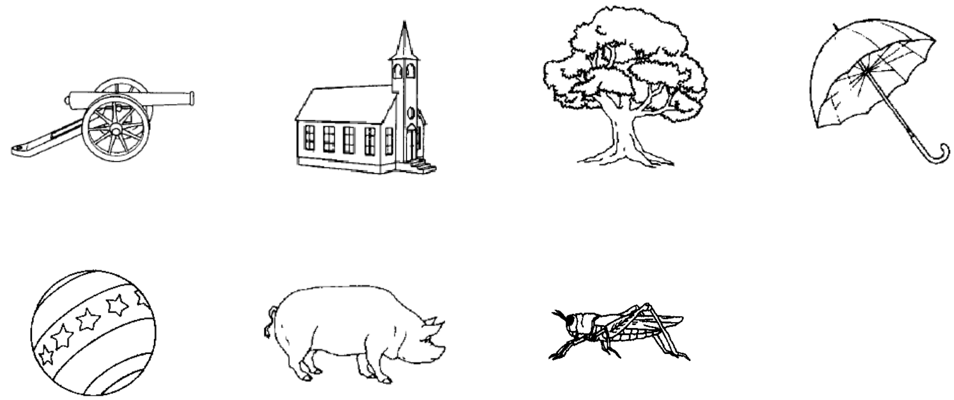
The items are listed in their order of presentation in the experiment.

List of images presented in chapter 6

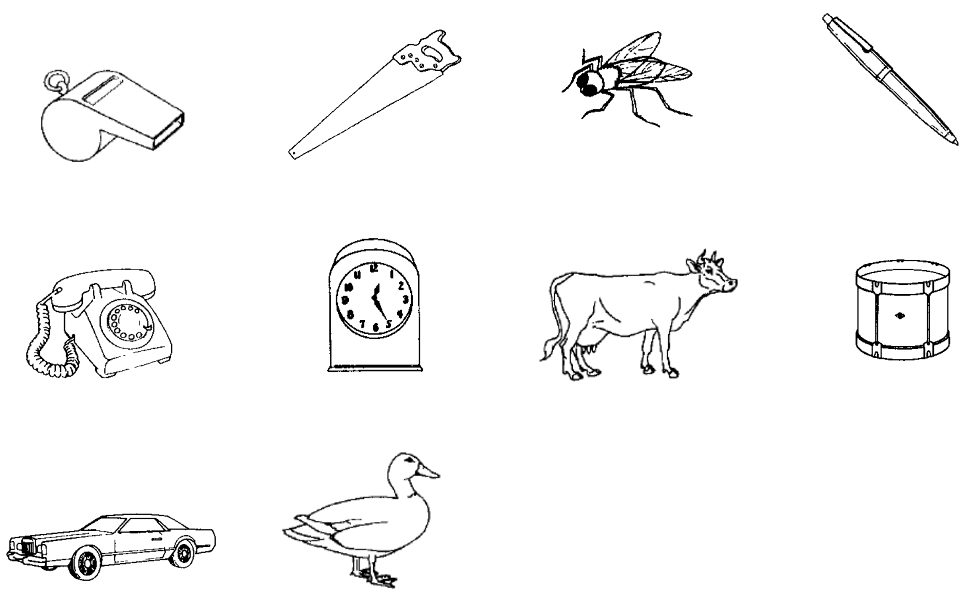


List of images presented in chapter 7

Maze 1

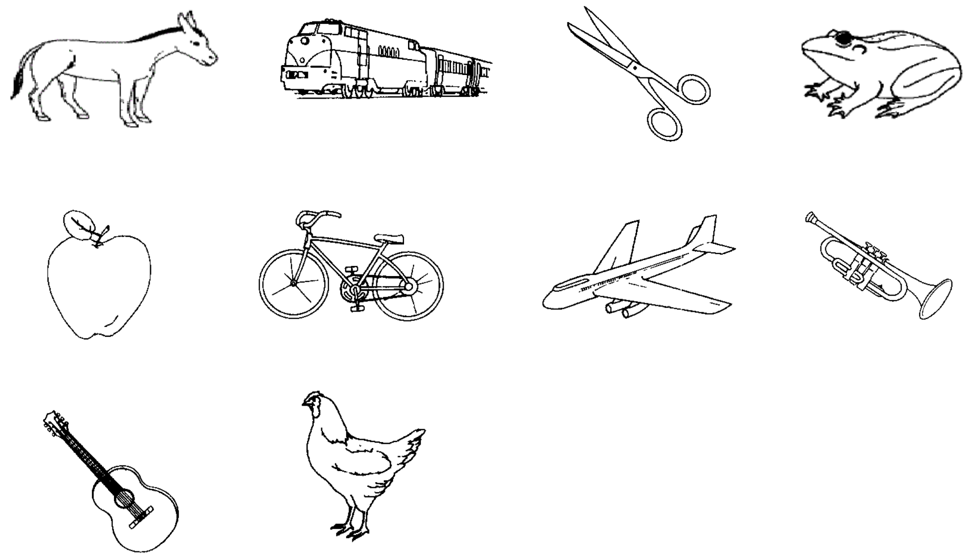


Maze 2

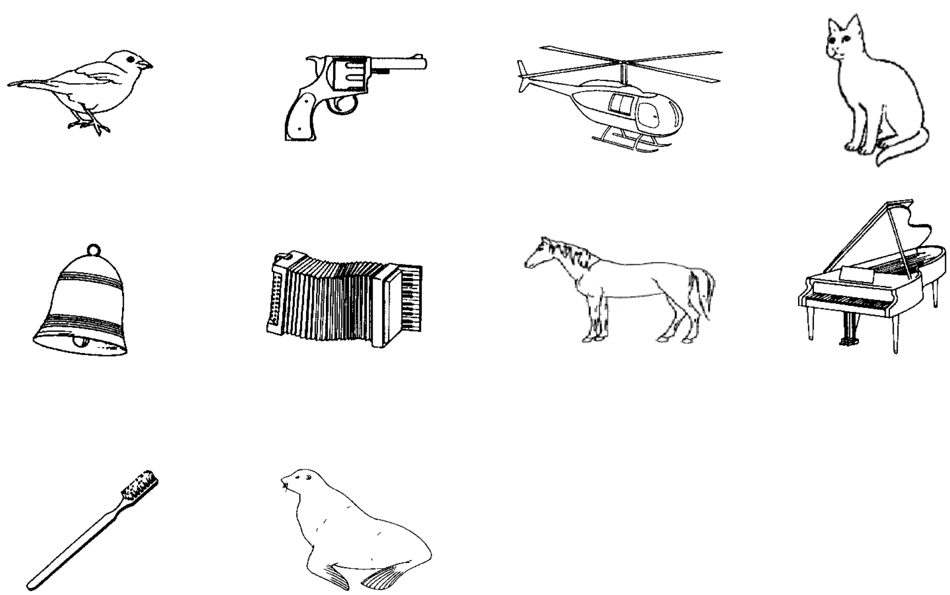


List of images presented in chapter 7 (continued)

Maze 3



Maze 4



Samenvatting

Interactie met onze omgeving is een van de meest fundamentele activiteiten in ons leven. Een belangrijk element van deze interactie is onze verplaatsing door de ruimte. Hier kan een onderscheid gemaakt worden tussen simpele beweging, dat wil zeggen de fysieke verplaatsing, en navigatie, een cognitief proces met betrekking tot het plannen en volgen van routes om van de huidige locatie bij een doel te komen.
Om succesvol te navigeren zijn de volgende vragen relevant: waar ben ik? waar wil ik heen? en hoe kom ik daar? Om deze vragen te beantwoorden maken we gebruik van onze waarneming en ons geheugen. Het waarnemen van een omgeving die we eerder bezocht hebben kan ons helpen om te bepalen waar we zijn, of we nog steeds een geplande route volgen en of we al op de gewenste doellocatie zijn aangekomen.
Navigatie kan plaatsvinden in zowel fysieke als virtuele werelden. Virtuele werelden zijn veelal drie-dimensionale simulaties van complexe (fysieke) omgevingen en kunnen tegenwoordig met behulp van computers eenvoudig gerealiseerd worden. Ze worden veelvuldig gebruikt voor vermaak (i.e. computerspelletjes), maar ook voor productontwerp en voor trainingsdoeleinden. De voordelen van trainen in virtuele omgevingen zijn kosten (vliegen in een vluchtsimulator is bijvoorbeeld goedkoper dan vliegen in een vliegtuig), veiligheid em flexibiliteit (omgevingsfactoren, zoals het weer, kunnen bijvoorbeeld met een druk op de knop gewijzigd worden).
Virtuele omgevingen verschillen echter van fysieke omgevingen. Virtuele omgevingen worden vaak op een computerscherm getoond en zijn meestal geheel visueel van aard. Het computerscherm heeft een beperkte resolutie en de virtuele omgeving heeft minder detail dan een fysieke omgeving. Ook verschilt de besturing vrijwel altijd ten opzichte van een fysieke wereld. Door deze verschillen is er minder informatie beschikbaar in virtuele dan in fysieke omgevingen.
Alhoewel mensen virtuele en fysieke omgevingen op een soortgelijke manier verkennen, ervaren ze, door de beperkte beschikbaarheid van informatie, meer problemen bij navigatie in virtuele omgevingen. Zo kost het meer tijd om de omgeving te leren kennen en verdwalen mensen vaker. Een mogelijke oplossing voor dit probleem is het presenteren van extra informatie. Deze informatie hoeft niet noodzakelijk visueel gepresenteerd te worden, maar kan ook via andere zintuigen aangeboden worden. Eerder onderzoek heeft aangetoond dat het aanbieden van informatie in meerdere zintuigen tegelijkertijd (multisensorische presentatie) gepaard gaat met prestatievoordelen. Zo reageren we sneller op en onthouden we meer informatie die in meerdere zintuigen tegelijkertijd wordt aangeboden.
Het doel van dit proefschrift is de mogelijke voordelen van de presentatie van visuele, auditieve en tactiele informatie op navigatie te onderzoeken. Echter, omdat navigatie berust op onze waarneming en geheugen, zijn multisensorische effecten op die processen zeer relevant. Een (groot) deel van dit proefschrift is dan ook gewijd aan het onderzoeken van die effecten op onze waarneming en geheugen.

Waarneming

In de eerste drie studies van dit proefschrift wordt het effect van multisensorische presentatie op de waarneming bekeken. In deze studies werden proefpersonen gevraagd korte reeksen van flitsen, piepjes en tikjes te tellen.

Hoofdstuk 2 beschrijft een experiment waarin we onderzoeken hoe het aanbieden van tikjes op de vinger het tellen van flitsjes beïnvloedt. Met behulp van electroencephalografie (EEG) laten we zien dat de activiteit in de visuele cortex, het hersengebied waar de flitsjes verwerkt worden, vlak na het aanbieden van een reeks flitsjes en tikjes verschilt wanneer de flitsjes en tikjes wel of niet tegelijkertijd werden aangeboden. Dit betekent dat er al vroeg in de informatieverwerking interacties optreden tussen de zintuigen. Dit kan een rol spelen in latere processen, zoals navigatie in virtuele omgevingen.

In hoofdstuk 3 en 4 vergelijken we de prestatie in het tellen van reeksen die aangeboden werden aan een of meerdere zintuigen (unisensorisch of multisensorisch). Uit hoofdstuk 3 blijkt dat er bij het tellen van multisensorische reeksen (bestaande uit flitjes, piepjes en tikjes) minder (grote) fouten worden gemaakt dan bij het tellen van unisensorische reeksen. In hoofdstuk vier laten we zien dat het voordeel van het aanbieden van multisensorische reeksen (bestaande uit flitsjes en piepjes) afhangt van de mogelijke lengte van de reeksen. De voordelen zijn veel kleiner als de reeksen bestaan uit maximaal drie flitsen en piepjes dan wanneer ze ze bestaan uit een tot tien flitsen en piepjes. Over het algemeen kunnen we stellen dat het voordeel van multisensorische presentatie groter wordt naarmate er meer fouten worden gemaakt in unisensorische reeksen. Voor virtuele omgevingen betekent dit dat multisensorische effecten op de waarneming alleen zullen optreden als de unisensorische componenten moeilijk waarneembaar zijn.

Geheugen

Het is reeds bekend dat het audiovisuele presentatie het opslaan van de identiteiten van objecten verbeterd. In hoofdstuk 5 onderzoeken we of tactiel-visuele presentatie het opslaan en het ophalen van zowel identiteiten als locaties van objecten verbeterd. Nadat proefpersonen vier abstracte objecten hadden verkend, moesten zij voor acht objecten aangeven of zij en waar zij die eerder hadden waargenomen. Uit de resultaten blijkt dat tactiel-visuele presentatie tijdens het opslaan of het ophalen de prestatie verbeterd ten opzichte van visuele presentatie, maar niet ten opzichte van tactiele presentatie. Dit geldt zowel voor het onthouden van identiteiten als voor het onthouden van locaties. De multisensorische effecten op geheugenprestatie zijn dus veel breder dan bekend was.

Hoofdstuk 6 beschrijft de invloed van de betekenis van objecten op de multisensorische geheugeneffecten. Deze studie werd uitgevoerd omdat de effecten die gerapporteerd zijn in hoofdstuk 5 kleiner zijn dan die in andere multisensorische geheugenstudies. In deze studie werd een serie betekenisvolle objecten (zoals plaatjes en geluidjes van dieren) en een serie betekenisloze objecten (Morse codes) aangeboden. Na de aanbieding van elke serie noteerden proefpersonen alle objecten die ze hadden onthouden. Uit de resultaten blijkt dat multisensorische presentatie alleen de geheugenprestatie verbeterd voor betekenisvolle objecten maar niet voor betekenisloze objecten. Dit laat zien dat het niet alleen belangrijk is objecten tegelijkertijd aan te bieden, maar dat ze ook (dezelfde) betekenis moeten hebben.

Navigatie
Tot slot onderzoeken we in hoofdstuk 7 of het presenteren van multisensorische oriëntatiepunten navigatie verbeterd. Proefpersonen verkenden een aantal virtuele doolhoven met visuele, auditieve of audiovisuele oriëntatiepunten. In vergelijking met verkenning met auditieve en visuele oriëntatiepunten resulteerde verkenning met audiovisuele oriëntatiepunten in een betere representatie van de virtuele omgeving. Vervolgens bleek dat proefpersonen sneller hun weg vonden in een wereld die ze verkend hadden met audiovisuele oriëntatiepunten dan met auditieve of visuele oriëntatiepunten. Het is aannemelijk dat een betere representatie van de virtuele wereld leidt tot efficiënter navigatiegedrag.

Conclusie
Alle studies beschreven in dit proefschrift laten zien dat het tegelijkertijd presenteren van informatie aan meerdere zintuigen de prestatie verbeterd. De verbeteringen variëren in sterkte van 10 tot 75% en doen zich voor bij de waarneming, het geheugen en bij navigatie.
De kennis verworven in deze studies draagt bij aan het in kaart brengen van de interacties tussen de zintuigen en geeft meer inzicht in de onderliggende mechanisme. De resultaten impliceren dat multisensorische presentatie (sommige van) de huidige problemen met navigate in virtuele werelden kan verlichten of oplossen. We verwachten dat multisensorische presentatie een navigator helpt informatie over zijn omgeving te onthouden, wat hem in staat stelt om efficiënter te navigeren in elke (virtuele) omgeving.

Dankwoord

Veel mensen hebben mij direct of indirect geholpen bij het tot stand komen van dit proefschrift. Graag wil ik hierbij een aantal van hen bedanken.

Allereerst wil ik mijn begeleiders bedanken. Peter en Jan zijn enorm betrokken geweest bij dit project. De vele gesprekken die ik met hen gevoerd heb zijn bijzonder enthousiasmerend en inspirerend geweest en brachten mij vaak tot nieuwe inzichten. Ik bewonder de scherpte waarmee zij telkens weer naar mijn stukken keken.

De leescomissie, bestaande uit Astrid Kappers, Albert Postma, Marc Ernst en Marta Olivetti Belardinelli wil ik graag danken voor het lezen van mijn proefschrift en het leveren van feedback.

Tijdens mijn promotie ben ik zowel werkzaam geweest op de Universiteit Utrecht alsmede bij TNO in Soesterberg. In het begin van mijn AiO-periode, toen er experimenten opgezet en uitgevoerd werden, ben ik veel bij TNO geweest. Later heb ik vaker op de Universiteit gezeten om over deze experimenten te rapporteren.

Bij TNO zou ik graag Antoon willen bedanken voor de hulp bij het opzetten van de experimenten en voor de leuke praatjes die wij maakten. Ook wil ik graag de collega's van de afdeling Perceptual & Cognitive Systems bedanken. In het bijzonder noem ik Anne-Marie en Maarten voor de feedback die zij gaven op mijn papers. Ook dank ik Ksander en Bruno en de andere AiO's bij TNO voor de pret die wij tijdens ons werk hebben gehad.

Bij de Universiteit wil ik graag bedanken Marc, Rita en Piet. Daarnaast zijn Marieke en Benny een grote steun en toeverlaat voor mij geweest, vooral bij de laatste loodjes van mijn proefschrift. Ook wil ik graag de Physics of Men groep bedanken voor de groepsbesprekingen die ik heb mogen bijwonen. Ik ging hier altijd met veel plezier naartoe en het heeft mij veel inzicht verschaft.

Tijdens mijn promotie heb ik verschillende studenten en stagiaires mogen begeleiden. In het bijzonder wil ik Hanna, Ksander, Rianne en Maya bedanken voor hun werk en gezelligheid. Ook gaat mijn dank uit naar hun begeleiders Stella en Albert.

Gedurende afgelopen vier jaar ben ik niet alleen maar bezig geweest met werk. Daarom wil ik mijn vrienden bedanken voor alle welkome afleiding zoals etentjes, spelletjes en andere gezellige activitieiten. Mijn familie, Bert, Atie, Aline, Eric, Freek, Hanneke, Jip, Inge, Kyra, Kees, Wilma, Mathijs en Franscisca wil ik bedanken voor hun steun en warmte. Tot slot wil ik Melanie bedanken voor haar liefde en geduld gedurende mijn promotie.

Curriculum Vitae

Tom Philippi was born in Rhenen, the Netherlands, on June 19, 1983. In 2001 he completed his secondary education at the Christelijk Lyceum in Veenendaal. He then studied Cognitive Artificial Intelligence at the Utrecht University, obtaining his Master's degree in 2007.

Immediately thereafter he obtained a PhD position at the department of Information and Computing Sciences at the Utrecht University. As part of this position he was also employed at the Department Perceptual and Cognitive Systems at TNO in Soesterberg, where he had previously conducted an internship. The topic of his research was multisensory interactions in perception, memory and navigation. The research was supervised by prof. dr. Peter Werkhoven and dr. Jan van Erp and resulted in the present thesis.

During his work as a PhD student, Tom was a member of the organizing committee of the EuroHaptics 2010 conference, which was held 7-9 July, 2010 in Amsterdam, The Netherlands.