Feature Integration across Multimodal Perception and Action

Sharon Shafir Zmigrod
Feature Integration across Multimodal Perception and Action

Proefschrift
ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus, Prof. mr. P. F. Van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 9 september 2010
klokke 13:45 uur

door

Sharon Shafir Zmigrod
Geboren te Petha-Tiqwa, Israel
in 1967
Promotiecommissie:

Promotor: Prof. dr. B. Hommel

Overige leden: Prof. dr. H. Swaab-Barneveld
Prof. dr. A. Cohen, Hebrew University, Israel.
Dr. G. Wolters
Dr. L. S. Colzato
“By sight I have the ideas of light and colours, with their several degrees and variations. By touch I perceive hard and soft, heat and cold, motion and resistance, and of all these more and less either as to quantity or degree. Smelling furnishes me with odours; the palate with tastes; and hearing conveys sounds to the mind in all their variety of tone and composition. And as several of these are observed to accompany each other, they come to be marked by one name, and so to be reputed as one thing. Thus, for example, a certain colour, taste, smell, figure, and consistence having been observed to go together, are accounted one distinct thing, signified by the name apple;” George Berkeley (1843)
# Table of Contents

Chapter 1: Introduction  
Chapter 2: Auditory event files: Integrating auditory perception and action planning.  
Chapter 3: Intermodal event files: Integrating features across vision, audition, taction, and action.  
Chapter 4: Temporal dynamics of unimodal and multimodal feature binding.  
Chapter 5: Cognitive flexibility and control in children with autistic spectrum disorder.  
Chapter 7: General discussion and conclusions  

References  

Summary in Dutch (samenvatting)  

Acknowledgments  

Curriculum Vita
Chapter 1

Introduction
Chapter 1 - Introduction

Introduction

One of the most remarkable aspects of multimodal perception is its coherence. Our conscious perception is unified at any given moment, although we acquire information from diverse channels with distinct transduction mechanisms, and process it in different cortical areas not necessarily at the same time and pace. For instance, a simple event such as eating a sandwich requires integration of the visual attributes such as the shapes and the colors of the ingredients; tactile attributes such as the sandwich’s texture and the degree of hotness, not forgetting the chemical attributes such as the smell and the taste; along with the action of chewing that might produce a sound. The problem of how the brain integrates the different types of information, which are processed in distinct cortical regions to a unified event, is referred to in literature as the binding problem (Triesman, 1996).

The study of the binding problem spans across many disciplines, ranging from understanding the unity of consciousness in philosophy, examining integration processes in cognitive science, and exploring the neural mechanisms in cognitive neuroscience. In terms of cognitive science, the binding problem focuses on the modularity of the brain. Each sensory modality processes its sensory information independently in specialized areas. For example, visual information is processed in the occipital lobe; auditory information in the temporal lobe; somatosensory information such as touch, pain, temperature, and proprioception in the parietal lobe, etc. (Gazzaniga, Ivry, & Mangun, 2002). Furthermore, each sensory modality system also processes different features in a specialized area, for instance, in the visual system, color is processed in V4 area and motion in MT (Zeki, 1993), similar findings were found in the auditory system (Lee & Winer, 2005) and in the parietal lobe, the somatosensory area (Culham & Kanwisher, 2001). In addition, responding and perceiving to an event requires planning and execution of actions, which are also processed in distinct areas (frontal lobe). Thus, within the brain the information about a specific event
is mostly distributed. At some point, the brain should construct some form of integrated representation for control and coherent perception, namely it needs to solve the *binding problem*.

**Feature integration - background**

One of the first and most influential theories in this domain was the Feature Integration Theory of attention (FIT), developed by Treisman and Gelade (1980). The theory posits that visual features (such as color, orientation, brightness, etc.) of an object are processed in parallel in separate feature maps and are later integrated through spatial attention or top down processes. Evidence for it comes from *object reviewing paradigm* (Kahneman, Treisman, & Gibbs, 1992), a visual task which measures performance of detecting a target letter on various moving objects. Better performance was achieved when the same letter appeared as part of the same object, an *object-specific preview benefit* that was taken to imply identity-location binding. Further research in this domain revealed that *object file* representation depends considerably on spatiotemporal information (Mitroff & Alvarez, 2007) and may persist for at least 8 sec (Noles, Scholl, & Mitroff, 2005). Additional support for the FIT theory comes from a study in the auditory domain by Hall, Pastore, Acker, & Huang (2000) where conjunctions of pitch and timbre were presented in different lateralized positions. The results demonstrated more frequent illusory conjunctions when pitch and timbre features were separately presented, suggesting that the auditory system binds its features with reference to their location, similarly to Feature Integration Theory in the visual domain.
Chapter 1 - Introduction

Multimodal integration

The focus on a single modality in feature integration dominated the field for some time. However, everyday events are not limited to a single modality, but rather they are multimodal in nature and should be examined as such. Exploring integration between the various modalities are more complex due to the differences in the physical attributes (such as properties of light, sound and touch propagation), the transduction mechanisms (reaching the brain at different points of time), the processing time (quicker for the auditory system than the visual), and the different cortical areas in which each sensory is processed. One of the methods that dominated the multimodal perception field was using conflict situations where two modalities receive incongruent information, creating different sorts of illusions. For instance, a classic example is the McGurk effect; in this effect, an auditory sound /ba/ paired with a visual lip movement associated with /ga/ often produces the percept /da/ (McGurk & MacDonald, 1976). Another audio-visual example is the ventriloquism effect; there localization of the sound source is shifted after exposure to a simultaneously auditory and visual stimulus but at disparate location (e.g., Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001). In the visual-tactile domain, it has been shown that an irrelevant visual distracter has influence on perceiving a tactile relevant stimulus, and is modulated by the spatial location of the stimuli (Spence, Pavani, & Driver, 2004). Additionally, studies demonstrated touch-induced visual illusion, where participants perceived one flash as multiple flashes when it was accompanied by more than one tap (Kunde & Kiesel, 2006; Violentyev, Shimojo, & Shams, 2005). Likewise, in the auditory-tactile domain, it has been shown that the auditory stimuli can alter the tactile perception, participants who received a single touch accompanied by more than one sound, perceived it as multiple touches (Hötting & Röder, 2004). These examples delineate the existence of cross-talk between different modalities. Yet, to understand coherent perception of
multimodal events, we need to investigate the processes in which it is achieved, that is the binding processes among various modalities.

**Feature integration across perception and action**

The traditional approaches in cognitive psychology differentiated between perception and action as two separate entities; however, there is ample evidence which supports a close and complicated relationship between these domains (for review see Noë, 2004). For example, a simple event such as looking at an apple requires the action of moving one’s head and eyes (saccades), even without considering eating it. Thus, perceiving an event is almost always accompanied by action. Hommel (1998, 2004) demonstrated that *object file* representations may contain action related information and are not purely perceptual, by designing a task based on the object reviewing paradigm (Kahneman, et al, 1992), which couple not only perceptual features but also perceptual features and the response features. In this task (event file task) two stimuli (combinations of two to three perceptual features) and two responses are presented. Each trial (see Figure 1.1) starts with the presentation of a response cue for the first response (R1), which has to be carried out after the presentation of the first stimulus (S1). The second stimulus (S2) is composed from the same, partly the same or totally different perceptual features than S1. The participants have to respond (R2) to one of the values of S2’s perceptual features, with the same or different response as R1. In this way, performance of the second event (S2 + R2) is affected by the preceding event (S1+R1). The general findings from such a task indicate costs (in terms of reaction time and accuracy) associated with repetition of some of the features but not all, either perceptual or response features, similar to the *object-specific preview benefit*.

Hommel (1998; 2004) has proposed that feature integration is not just perceptual but a general phenomenon that crosses domains such as perception and
action, and provided empirical evidence that features from the visual domain and action are integrated into an episodic representation (so-called event files: Hommel, 1998). The event file paradigm (see Figure 1.1) was used in many studies to explore further principles and constraints regarding the creation, maintenance (updating) and revision of such episodic representations (Hommel 2005, 2007b, 2009; Hommel & Colzato, 2004), providing essential information for understanding the underlying mechanisms of binding.

Figure 1.1. Sequence of events in the event file task. A visual response cue signals a left or right response (R1) that should be delayed until presentation of the first stimulus S1 (S1 is used as a detection signal for R1). The second stimulus S2 appears 500 ms after responding to S1. S2 signals R2, a speeded left or right response according to one of the values of S2.

Mechanisms of bindings

The implementation of multimodal binding mechanisms in a distributed brain requires an integration of diverse kinds of information from dozens of separate cortical areas. How can a neural mechanism perform such a complicated task of perceiving and acting in a coherent manner? First, it must be a dynamic system which can handle an enormous amount of feature combinations as well as novel combinations; second, it should make it possible for distinct features (i.e.
the color blue) to be bound to diverse objects or events simultaneously; third, it ought to enable connections between disperse anatomical regions that process different feature codes, for instance, posterior areas such as sensory areas, and more frontal areas such as planning, acting and control areas. Additionally, as it operates in a very dynamic environment, it should be flexible, yet sustainable and persistent for some time. Last, it must be robust as it operates as a core process in the brain. An early attempt to explain such a mechanism was by a convergence mechanism (Barlow, 1972), in which specialized high level neurons can detect feature conjunctions. However, this creates the so-called ‘combinatorial explosion’ problem, which requires too many units to cover all possible combinations in a single modality, and this problem increases exponentially when more than one modality is concerned.

Another proposed mechanism through which binding can be achieved is temporal synchrony (see: Engel & Singer, 2001; Raffone & Wolters, 2001; von der Malsburg, 1981, 1999). The basic idea is that separate pieces of information belonging to an event can be bound together by synchronizing their spiking rate. Thus, neurons, which fire in the same rhythm, represent the same event. This mechanism can overcome many issues brought about by a convergence mechanism. There is a growing body of empirical evidence, both in human and animals that support temporal synchrony as an integration mechanism in different modalities. For instance, studies found neural synchronization in the gamma range (~30-100 Hz) in the visual areas (Engel, Konig, & Singer, 1991), auditory areas (deCharms & Merzenich, 1996; Joliot, Ribary, & Llinás, 1994), and somatosensory areas (Murthy & Fetz, 1992; Nicolelis, Baccala, Lin, & Chapin, 1995). Also, evidence for neural synchronization was found between different modalities and domains in the beta range (~12-20 Hz) such as, between visual and auditory areas (von Stein, Rappelsberger, Sarnthein, & Petsche, 1999), between visual and the motor areas (Roelfsema, Engel, Konig, & Singer, 1997), and between motor and somatosensory areas (Murthy & Fetz, 1992; 1996). Taking
together, these findings suggest the existence of local synchronized activity which might underlie feature integration and object representation (Tallon-Baudry & Bertrand, 1999). However, recent papers (Hommel & Colzato, 2009; VanRullen, 2009) suggested that more than one neural mechanism may be responsible for feature integration in the visual domain. Also, there are several computational models that support the conjunctive binding approach (O'Reilly, Busby, & Soto, 2003). Thus, there is still room for new and reconcile approaches as candidate mechanisms to the binding problem.

**Thesis question**

Feature binding is a core process in perception and action if not in many of the cognitive functions in the brain, due to the brain’s distributed architecture. Exploring the different aspects of the binding mechanisms may help us to gain a better understanding of this key component. Until now, the emphasis in perception and action domains was limited to unimodal integration and even more specifically to the visual domain. However, the binding problem is equally valid and central to multisensory perception. Moreover, some of the principles and constraints might be manifested through careful examination of effects between sensory modalities and domains. This thesis intended to look into effects, attributes, principles, and constraints of how the brain binds different features within and across modalities and domains. As Treisman (2003) stated, this mechanism is not introspectable and needs to be investigated through careful empirical studies. Thus, this work focused on behavioral data and provided an empirical evidence for integration effects, integration principles, and constraints concerning event file management.
Outline of thesis

The current thesis contains five chapters of empirical studies (chapters 2 - 6) and ends with a general discussion (chapter 7). It reflects a gradual inquest in order to reveal different aspects of the binding mechanism across multimodal perception and action.

The current chapter (1) introduces traditional and more contemporary theories and frameworks in the domain of feature integration and multimodal perception.

Chapter 2 explores the binding mechanisms within and across the auditory domain and action planning, and describes the temporal overlap principle as one involving in this mechanism. Additionally, the mediating role of attention in the integration processes is discussed.

Chapter 3 provides empirical evidence for multimodal feature integration in the visual, auditory, and tactile domains along with binding between those perceptual domains and action domain. The results of the multimodal experiments reveal the same types of interactions, as for unimodal feature combinations, yet the size of the interactions varies with the particular combination of features, suggesting that the salience of features and the temporal overlap between feature-code activations plays a mediating role. Thus, the findings here confirm that feature integration operates under general principles and crosses modalities and domains.

Chapter 4 explores the temporal dynamics of feature integration within a single sensory modality (the auditory system) and between modalities (such as visual and auditory modalities), as well as across perception and action. The findings show that integration effects decrease systematically with increasing time between the two stimuli and response events, and the decrease rate is comparable to unimodal and multimodal bindings, pointing to similar mechanisms.
Chapter 5 examines how control and flexibility are associated with event file maintenance, by comparing integration effects between populations with developmental disorders such as autistic spectrum disorder (ASD) children and typically developing children. The findings in this study demonstrate that ASD children are impaired in updating event file representations due to a lack of cognitive flexibility as was measured by an executive function task, suggesting a common ground between these cognitive functions presumably due to prefrontal dopaminergic hypoactivity.

The last empirical chapter (6) examines the relationship between feature binding and coherent perception (our consciousness) and argues that the binding processes are neither prerequisite nor consequence of unified perception. Thus, these findings break the symbiotic relationship between the two, and challenge some of the definitions in the literature.

Finally, chapter 7 discusses the empirical findings by describing the principles and the constraints and delineating these principles in a schematic model.
The research reported in the five empirical chapters is either published or submitted in international, peer-reviewed psychological journals. The reports appear in this thesis in their published/submitted form. A list of references is presented so as to acknowledge the contributions of the co-authors to these articles.


Chapter 2

Auditory Event Files: Integrating Auditory Perception and Action Planning

Abstract

The features of perceived objects are processed in distinct neural pathways, which call for mechanisms that integrate the distributed information into coherent representations (the “binding problem”). Recent studies of sequential effects demonstrate feature binding not only in perception but also across (visual) perception and action planning. We investigated whether comparable effects can be obtained in and across auditory perception and action. Results from two experiments revealed effects indicative of spontaneous integration of auditory features (pitch and loudness, pitch and location) as well as evidence for audio-manual stimulus-response integration. Even though integration takes place spontaneously, features related to task-relevant stimulus or response dimensions are more likely to be integrated. Moreover, integration seems to follow a temporal-overlap principle, with features coded close in time being more likely to be bound together. Taken altogether, the findings are consistent with the idea of episodic "event files" integrating perception and action plans.
Introduction

The perceived features of visual (Zeki & Bartels, 1999) and auditory (Kaas & Hackett, 1999; Lee & Winer, 2005; Wessinger et al., 2001) objects are processed in distinct neural pathways, which calls for processes that integrate this distributed information into coherent representations. This so-called “binding problem” and the mechanisms solving it have been studied extensively in recent years (e.g., Allport, Tipper, & Chmiel, 1985; Hall, Pastore, Acker, & Huang, 2000; Hommel, 2004; Treisman & Gelade, 1980). One of the leading theories in this field, Treisman’s Feature Integration Theory (FIT), holds that primary visual features are processed in parallel and represented in separate feature maps. Through spatial selection via a master map of locations an episodic representation is created: an “object file”, that is updated as the object changes and that can be addressed by location (Kahneman, Treisman, & Gibbs, 1992; Treisman, 1990; Treisman & Gelade, 1980).

Hommel (1998, 2004, 2005) extended Treisman’s “object file” concept to include not only stimulus features but also response-related feature information. A number of studies provided evidence for this extension. In these studies, participants carried out two responses in a row. First, they were cued by a response cue signaling the first response, which however was carried out only after a visual trigger stimulus was presented. After one second another visual stimulus appeared and the participants had to perform a binary-choice response to one of its features. As expected, main effects of stimulus-feature repetition were obtained. But more interestingly, stimulus and response repetition effects interacted: Repeating a stimulus feature sped up reaction time (RT) only if the response also repeated, whereas stimulus feature repetition slowed down RT if the response alternated. Apparently, stimulus features were bound to response features, so that repeating one retrieved the other. This created conflict in partial repetition trials, that is, when the retrieved stimulus or response feature did not
match the present one. Hence, facing a particular combination of stimulus and response features seems to create a multimodal “event file” (Hommel, 1998, 2004), which is retrieved if at least one of the features it includes is encountered again.

The existing theories in feature integration were largely based on experiments using visual information, but it makes sense to assume that feature integration takes place in auditory perception as well. The auditory system allows us to perceive events based on the sound produced by them. And yet, an acoustic event is commonly made up of several features among them pitch, timbre, loudness, and spatial position. Numerous studies looked into how these features are perceived; however, in everyday life we do not perceive features in isolation but, rather, coherent, integrated acoustic events. Given that these features are processed in different areas of the auditory cortex (Kaas & Hackett, 1999; Wessinger et al., 2001), there should be a mechanism that integrates the auditory features into a coherent acoustic perception. Indeed, there is preliminary evidence for the existence of auditory binding. For instance, Hall et al. (2000) examined auditory feature integration of spatially distributed musical tones by having participants search for either a cued conjunction of pitch and timbre or a single cued value (pitch or timbre) in arrays of simultaneous tones in different lateralized positions. Their finding revealed more frequent illusory conjunctions when pitch and timbre features were separately presented, suggesting that, similar to the visual system, the auditory system differentiates the auditory features from the sound field and then integrates them according to their source. The investigators concluded that the auditory system binds its features with reference to their location, just like the Feature Integration Theory (Treisman & Gelade, 1980) assumes for the visual system. Additionally, Leboe, Mondor, and Leboe (2006) who investigated different sources of auditory negative priming effects, found that repeated sounds in opposite locations were categorized slower than repeated sounds in the same location. In the inter-domains of auditory perception
and action, Mondor, Hurlburt, & Thorne (2003) found interactions between pitch- and response-repetition effects, which may indicate the integration of sound features and action.

Another important research question that has been addressed concerns the role of the attention in auditory feature binding. Previous studies have shown contradicting evidence. Hall et al. (2000) suggested that reliable integration of auditory features might require focused attention to avoid illusory feature conjunctions when multiple sounds exist. However, this suggestion is inconsistent with recent findings of Takegata and colleagues (2005). They conducted an EEG study in which participants performed a visual working memory task while ignoring a background of two sounds. The two sounds, varying in timbre and pitch, were played simultaneously. Regardless of the task load, the pitch-timbre combinations elicited similar amplitudes and latencies in the ERP component mismatch negativity (MMN). According to the investigators these results provided evidence that feature integration in the auditory modality can occur without focus of attention. In line with this view, Hommel (2005) demonstrated that even irrelevant visual stimuli may be bound to a response.

Although there is ample evidence for the existence of event files in and across visual perception and action planning, the event file concept has not been systematically applied to auditory perception and action planning. Only a few studies have examined the binding mechanism in the auditory modality, and there is contradictory evidence regarding the role of attention in this mechanism. The aim of the current study was to investigate feature binding mechanism in and across the auditory perception and action planning. More specifically, we addressed three research issues: whether evidence for feature integration in a standard "object file" can be observed for different auditory dimensions; whether evidence for stimulus-response integration effects can be obtained between the auditory modality and action planning; and whether these integration effects rely on, or are mediated by attention.
Figure 2.1: Sequence of events of the experiments. A response cue signaled a left or right mouse button click (R1) that was to be delayed until presentation of the S1. S2 appeared 1000 msec later. S2 signaled R2, a speeded left or right mouse button click according to the task.

Experiment 1

Experiment 1 was performed to determine whether auditory features are integrated into a coherent object representation and whether response-related features are also integrated with auditory features to produce an “event file” similar to Hommel’s (1998, 2004) findings in the visual domain. The task followed Hommel’s (1998) design, only that the stimuli were pure tone sounds. Participants were cued to prepare a response (left or right mouse button click), which they carried out (R1) after the first stimulus (S1). One second later the second sound (S2) was played and participants had to respond to the value of the relevant auditory feature by carrying out response R2 (left or right mouse button click) (see Figure 2.1).

The auditory features that were chosen for this experiment were pitch and loudness. Neuhoff, Kramer, and Wayand (2002) demonstrated that pitch and loudness have an interactive effect, that is, changes in one of these dimensions influenced the other. Based on these results and the “object file” concept, we
hypothesized that pitch and loudness features of S1 are integrated and are still bound when processing S2. If so, repeating the feature in one dimension should produce better performance if the feature in the other dimension is also repeated; whereas alternating the feature in one dimension should produce better performance if the feature in the other dimension is also alternated. In addition, we hypothesized that the features making up S1 are integrated with R1 and are still bound to it when responding to S2, based on the suggested event-file mechanism which posits that specific combination of stimulus and response creates episodic trace that is retrieved in case of any feature repetition (Hommel, 1998, 2004). If so, response to S2 should be better with a complete match or a complete mismatch between the previous response and a given auditory feature than with partial matches. Moreover, previous observation showed that pitch repetition interacts with response repetition (Mondor et al., 2003). To investigate the role of attention in auditory feature integration we manipulated the feature that was relevant for responding to S2. In one block of trials, only one of the two auditory features (pitch and loudness) was relevant, while in another block the other auditory feature was relevant. Task relevance of S2 features (and the amount of attention consequently devoted to them) has been shown to affect the size of integration-related effects with visual stimuli (e.g., Hommel, 1998), and we were interested to see whether it would also modify such effects with auditory stimuli.

Method

Participants

Fourteen participants were recruited by advertisement for this experiment and were paid or received a course credit for a 40 min session. Two participants were excluded from the analysis due to a high error rate (around chance level 50%) and very slow RT in the pitch task—reflecting their difficulty in identifying low vs. high pitch (see Neuhoff, Knight, & Wayand, 2002). The remaining 12
participants (4 male; mean age 23, range 18-38 years) reported not having any known hearing problem. The participants were naïve as to the purpose of the experiment.

**Apparatus and stimuli**

The experiment was controlled by a Targa Pentium 3, attached to a Targa TM 1769-A 17-inch monitor. Participants faced the monitor at a distance of about 60 cm. The loudspeakers were located on both sides of the screen (approximately 25 degrees) at a distance of 70 cm. The stimuli S1 and S2 were composed from two pure tones of 1000Hz and 3000Hz with duration of 50 msec and were presented at 65 dB SPL and 75 dB SPL. Visual response cues were presented in the middle of the screen (see Figure 2.1) with right or left arrow indicating a right and left response (R1), respectively. Responses were made by clicking on the left or the right mouse button with index and the middle finger of the dominant hand.

**Procedure and design**

The experiment was composed of two sessions: in one session pitch was the relevant dimension for the task and the subjects had to respond to whether pitch was high or low; in the other session loudness was the relevant dimension for the task and the subjects had to respond to whether loudness was high or low. The sessions were counterbalanced between subjects. Each session contained a practice block with 10 practice trials and an experimental block with 128 experimental trials. The order of the trials was randomized. Participants had to carry out two responses per trial: R1 was a simple reaction with left or right mouse click as indicated by the direction of an arrow in the response cue. It had to be carried out as soon as S1 appeared, regardless of its pitch or its loudness. R2 was a binary-choice reaction to S2. In the pitch-relevant session half of the participants responded to the high pitch (3000Hz) and the low pitch (1000Hz) by pressing on the left and right mouse button, respectively, while the other half received the opposite mapping. In the loudness-relevant task half of the
participants responded to the loud sound (75 dB SPL) and to the soft sound (65 dB SPL) by pressing on the left and right mouse button, respectively, while the other half received the opposite mapping. The participants were asked to respond as quickly and accurately as possible.

The sequence of events in each trial is shown in Figure 2.1. A response cue with a right or left arrow was visually presented for 1500 msec signaled response (R1) which was to be carried out after stimulus 1 was played. S2 was played one second after the response to S1, with the pitch (in the pitch session) or loudness (in the loudness session) signaling the second response (R2). In case of incorrect or absent responses an error message was presented. R2 speed (reaction time or RT) and accuracy (percentage of errors or PE) were analyzed for all trials with correct R1 responses as a function of session (pitch/loudness), repetition vs. alternation of the response, and repetition vs. alternation of the stimulus dimensions pitch and loudness.

Results

Trials with incorrect R1 responses (1.7%), as well as missing or anticipatory (RT<100 msec) R2 responses (0.7%) were excluded from analysis. The mean reaction time for R1 was 270 msec (SD=88). From the remaining data, mean RTs and PEs for R2 were analyzed as a function of the four variables: the task-relevant stimulus feature (loudness vs. pitch) or task for short, the relationship between the responses R1 and R2 (alternation vs. repetition), the relationship between S1 and S2 on the pitch dimension (alternation vs. repetition), and the relationship between S1 and S2 on the loudness dimension (alternation vs. repetition). ANOVAs were performed by using a four-way design for repeated measures. Table 2.1 provides an overview of the RT and PE means obtained for R2 performance.
Table 2.1. Means and standard errors of mean reaction times (RT in msec) and percentages of errors (PE) for responses to stimulus 2 (R2) as a function of the attended dimension, the relationship between the stimuli (repetition vs. alternation) and the relationship between the responses (repetition vs. alternation).

<table>
<thead>
<tr>
<th>Attended Dimension</th>
<th>Stimulus Feature</th>
<th>Response</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Repeated</td>
<td>Alternated</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>RT (SE)</td>
<td>PE (SE)</td>
<td>RT (SE)</td>
<td>PE (SE)</td>
</tr>
<tr>
<td>Loudness</td>
<td>Neither</td>
<td>553 (41)</td>
<td>8.8 (3.0)</td>
<td>472 (31)</td>
</tr>
<tr>
<td></td>
<td>Loudness</td>
<td>557 (24)</td>
<td>12.4 (2.8)</td>
<td>553 (36)</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td>553 (31)</td>
<td>11.0 (4.5)</td>
<td>517 (27)</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>486 (32)</td>
<td>6.3 (3.2)</td>
<td>541 (35)</td>
</tr>
<tr>
<td>Pitch</td>
<td>Neither</td>
<td>574 (41)</td>
<td>11.3 (2.2)</td>
<td>502 (39)</td>
</tr>
<tr>
<td></td>
<td>Loudness</td>
<td>564 (38)</td>
<td>14.6 (3.9)</td>
<td>521 (40)</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td>548 (42)</td>
<td>19.1 (5.3)</td>
<td>604 (38)</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>507 (42)</td>
<td>7.9 (2.7)</td>
<td>545 (45)</td>
</tr>
</tbody>
</table>

First we report less important theoretical findings; the analysis yielded a main effect of pitch in PEs, $F(1,11)=5.22, p<.05$, with higher error rates for pitch repetition than alternation. This effect was further modified by task $F(1,11)=8.54, p<.05$, indicating that it was more pronounced in the pitch task $F(1,11)=11.13, p<.01$ than loudness task $F<1$. Similarly, interaction between loudness and task in PEs was obtained, $F(1,11)=5.28, p<.05$, which was also more pronounced in the pitch task $F(1,11)=7.42, p<.05$ than loudness task $F<1$.

Second we address the stimulus-integration effect by examining the interactions between repetition vs. alternation of the stimulus features: there was an interaction between pitch repetition (vs. alternation) and loudness repetition, $F(1,11)=11.07, p<.01$, indicating that, with pitch repetition, performance was quicker if loudness was also repeated than if loudness was alternated; whereas,
with pitch alternation, performance was quicker if loudness alternated than if it was repeated (see Figure 2.2). This result provides support for auditory feature integration between pitch and loudness.

**Figure 2.2.** Reaction times in Experiment 1, as a function of repetition vs. alternation of pitch and loudness.
Third, we consider stimulus-response-integration effects by examining the interactions between repetition vs. alternation of the response and the stimulus features. There were interactions between response repetition and pitch repetition in RTs, \( F(1,11)=42.45, p<.0001, \) and PEs, \( F(1,11)=8.90, p<.05, \) showing that response repetition facilitates performance if the pitch repeats but impairs performance if the pitch alternates. Furthermore, there was an interaction between response repetition and loudness repetition in RTs, \( F(1,11)=5.14, p<.05, \) and PEs, \( F(1,11)=9.30, p<.05, \) showing that the responses were faster and more accurate for total repetition or total alternation of the response and the loudness than partial repetition. Additionally, a three-way interaction among task, response, and loudness in RTs, \( F(1,11)=6.63, p<.05 \) was obtained, indicating sensitivity to task-relevance feature in this stimulus-response effect. Separate ANOVAs confirmed that response only interacted significantly in RTs with loudness in the loudness task, \( F(1,11)=7.38, p<.05 \) and not in the pitch task \( F<1. \) These interactions show stimulus-response effects between the response and the auditory stimuli. In the case of loudness, it was modulated by task relevance (see Figure 2.3).
Figure 2.3. Reaction times in Experiment 1 for the repetition vs. alternation of relevant and irrelevant stimulus (pitch and loudness), as a function of response (repetition vs. alternation) in pitch and loudness task.
Discussion

Experiment 1 was successful in providing evidence for event file creation in auditory perception and action planning. It demonstrated the spontaneous integration of pitch and loudness even when only one of the dimensions was the task relevant and the other could be ignored. In addition, we observed stimulus-response integration effects for pitch and loudness, which were more pronounced for the task relevant feature. This is in line with findings from visual studies, where integration was also spontaneous (i.e., occurred even if unnecessary for the task) but was mediated by task relevance of the feature dimensions (see Hommel, 2004, for an overview).

Our findings seem consistent with a recent auditory study of Mondor and Leboe (2008). These authors observed that the impact of pitch repetition on tone-detection performance depends on response repetition—which seems to fit with our present stimulus-response-integration effects. In particular, they found pitch-repetition benefits if both the prime and the probe tone were to be detected and thus accompanied by the same response, and pitch-repetition costs if the prime was to be ignored and thus not accompanied by a response. This outcome pattern bears similarities with our findings: good performance if both pitch and response repeat or both alternate, but bad performance is one repeats but not the other. However, Mondor and Leboe (2008) manipulated the response requirements between participants, which may have induced different attentional sets and strategies in the two tasks. For instance, ignoring primes in a detection task may lead to “inhibition of return” (Posner & Cohen, 1984), which may explain stimulus-repetition costs without referring to response requirements—and indeed, ignoring the prime and omitting a response to it lead to a 70-msec increase of reaction time. Accordingly, it is not clear whether the observations of Mondor and Leboe reflect the same mechanisms that underlie the pitch-by-response interactions obtained in the present study.
Our findings reveal an interesting dissociation between the integration of stimulus features and the integration of stimulus and response features—a dissociation in which attention induced by task relevance plays a major role. Stimulus-response integration seems to be mainly restricted to the stimulus features that are task relevant: pitch in the pitch task and loudness in the loudness task. In contrast, different features of the same stimulus seem to be integrated irrespective of task relevance, as evidenced by the reliable interaction between pitch and loudness under conditions that rendered only one of them relevant at any given time, one possible explanation might be that the physical attributes of the features influence one another, i.e. loudness is known to be affected by frequency and pitch by intensity. It is also interesting to see that, in stimulus-response integration, the effect of task relevance was more effective in excluding irrelevant loudness information than irrelevant pitch information. In other words, in the current study loudness was more sensitive to task relevance than pitch.

We think that all these aspects of our findings point to the same integration principle: features of events (whether they refer to stimuli or responses) are integrated to the degree that the activations of their codes overlap in time. This principle underlies the concept of conditioning (Pavlov, 1927) and seems crucial for the hippocampal integration of episodic stimulus and action events (Bangasser, Waxler, Santollo, & Shors, 2006). First, consider the respective roles that this principle plays in the integration of stimulus features versus the integration of stimulus and response features. As indicated in panel A of Figure 2.4, the activations of stimulus feature codes are likely to overlap in time even if they are peaking at different time points, that is, even if stimulus features are registered asynchronously. Accordingly, they are likely to be bound to each other, thus producing a partial-overlap cost. However, the earlier a feature is coded the earlier its code decays, suggesting that quickly coded features are less likely to overlap in time with response code activation. In our study, we found that RTs were faster in the loudness task than the pitch task (see Figure 2.3)
probably due to the greater saliency of loudness and/or better discriminability of the loudness values we chose, suggesting that in this experiment loudness was coded faster than pitch (FOOTNOTE 1). With respect to the temporal relations depicted in Figure 2.4, this implies that response code activation started earlier in our experiment in the loudness task than it did in the pitch task. On top of that, there is evidence that loudness codes decay faster than pitch codes do (Clement, Demany, & Semal, 1999), which would further work against the integration of loudness and response. We can thus conclude that the code-overlap principle accounts for both the observation that task relevance did not affect stimulus integration and the finding that it did affect stimulus-response integration.

Making a feature dimension relevant to a task is likely to increase the weights (or gain) of that dimension's codes (Bundesen, 1990; Found & Müller, 1996; Hommel, Müsseler, Aschersleben, & Prinz, 2001), which again may result in stronger and/or more enduring activation (see panel B in Figure 2.4). This means that task-relevant features induce activations that are more likely to overlap with the response activation. As a consequence, task relevant features should be more likely to be integrated with the response than task irrelevant features, just as we observed in Experiment 1.
**Figure 2.4.** Sketch of the hypothetical activation functions of stimulus codes. A. In our experiment loudness was coded faster than pitch was, so that the activation of pitch codes (even as the irrelevant dimension) is more likely to overlap with response-code activation. B. Task relevance of a given feature increases the duration of code activation, so that even codes that are activated early in time are now overlapping with response code activation.
Chapter 2 - Auditory Event Files

**Experiment 2**

Experiment 1 suggests that pitch and loudness are spontaneously integrated both with each other and with the response, at least if the given feature is task relevant. Experiment 2 investigated whether these observations can be extended to stimulus location. Many authors have emphasized the possibly crucial role of stimulus location in feature integration (in vision: Treisman & Gelade, 1980; in audition: Hall et al., 2000; Leboe, Mondor, & Leboe, 2006).

On the one hand, this could mean that spatial location is so important for feature integration that it does not matter whether location information is nominally relevant or irrelevant for a given task. This would still be consistent with the feature-overlap principle, assuming that location features are strongly weighted irrespective of the task, but it would imply that the proposed relationship between task relevance and weighting does not apply to location. On the other hand, however, it is true that many tasks that are taken to demonstrate the crucial role of location have used spatial responses. Assuming that responses are represented, prepared, and planned in terms of their perceptual features (Hommel, 1996; Hommel et al., 2001), it is possible that defining a response set in terms of spatial features (e.g., by characterizing responses as "left" and "right") attracts attention to the spatial dimension(s) and, thus, induces a stronger weighting of spatial codes. Indeed, Fagioli, Hommel, and Schubotz (2007) found evidence that preparing for particular types of actions (grasping versus pointing) attracts attention to the features that are relevant for defining these actions (size versus location). Along the same lines, Hommel (2007b) observed that the integration of visual stimulus location and the response is much more pronounced when the response alternatives are spatially defined (left versus right) than when they are not (pressing a key once versus twice). Hence, it is possible that the previous findings of integration of (nominally) irrelevant location information and the response do not so much reflect a central role of stimulus location in feature
integration but, rather, the fact that defining responses spatially makes location task relevant.

The aim of Experiment 2 was to examine this possible interpretation of the role of location information, apart from studying the integration-related effects of the auditory location as such. We did so by manipulating the pitch and location of auditory stimuli and by using two different types of response sets. One set was spatially defined, just as in Experiment 1, and the other consisted of a non-spatial Go/No-Go response. We expected to replicate the findings from Experiment 1 with regard to pitch and to obtain comparable findings for location. However, the location-related findings should vary with the response set, with the spatial set producing stronger integration of location codes than the non-spatial set.

**Method**

**Participants**

Thirty participants were recruited by advertisement for this experiment and were paid or received a course credit for 40 minutes session. One participant was excluded from the analysis due to a high PE (around chance level 50%) and very slow RT in the pitch task. The remaining 29 students (3 male; mean age 22, range 18-34 years) reported not having any known hearing problem. They were randomly assigned to two groups, a spatial response set group (N=14) and a non-spatial response set group (N=15).

**Procedure and design**

The procedure was as in Experiment 1, with the following exceptions. The loudspeakers were placed at an upper and lower position at 45 degree from the center of the screen. The stimuli S1 and S2 were composed from two pure tones of 1000Hz and 3000Hz with duration of 50 msec, presented at approximately 70 dB SPL. The experiment was composed of two sessions: in one session pitch was relevant for responding to S2; in the other session location was relevant to S2.
requiring a response to the top versus bottom location. The sessions were counterbalanced between subjects. Each task contained a practice block with 15 practice trials and an experimental block with 96 experimental trials. The order of the trials was randomized.

The spatial response set group saw a left or right arrow indicating a left and right mouse click, respectively; responses to S1 and to S2 were made like in Experiment 1. The non-spatial response set group saw the word GO or NO GO, indicating whether to emit or withhold the response, respectively. Responses on the GO trials were made by clicking on the left mouse button; the NO GO trials for S1 lasted 500 msec.

**Results and Discussion**

Trials with incorrect R1 responses (1%) as well as missing or anticipatory R2 responses (RT<100 msec) (0.1%) were excluded from analysis. The mean reaction times for R1 were 330 msec (SD=78) for the spatial response-set group and 341 msec (SD=114) for the non-spatial response-set group. From the remaining data, mean RTs and PEs for R2 were analyzed as a function of the five variables: the task (pitch vs. location as relevant S2 feature); the relationship (repetition vs. alternation) between S1 and S2 with regard to pitch and location, the relationship (repetition vs. alternation) between responses R1 and R2; and the response set (spatial vs. non-spatial) (see Table 2.2 for mean RTs and PEs). ANOVAs were performed by using a mixed design with repeated measures on four variables and with response set as between group variable, see Table 2.3 for the outcomes.
**Table 2.2.** Experiment 2: Means and standard errors of mean reaction times (RT in msec) and percentages of errors (PE) for responses to stimulus 2 (R2) as a function of the response set (spatial or non-spatial), attended dimension, the relationship between the stimuli (repetition vs. alternation), the relationship between the responses (repetition vs. alternation).

<table>
<thead>
<tr>
<th>Response Set</th>
<th>Attended Dimension</th>
<th>Stimulus Feature</th>
<th>Repeated</th>
<th>Alternated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Repeated RT (SE)</td>
<td>PE (SE)</td>
<td>RT (SE)</td>
</tr>
<tr>
<td>Spatial</td>
<td>Location</td>
<td>Neither 496 (33)</td>
<td>12.8 (3.1)</td>
<td>453 (23)</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Location 534 (28)</td>
<td>17.1 (2.8)</td>
<td>542 (24)</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td>Pitch 506 (30)</td>
<td>8.0 (2.3)</td>
<td>496 (25)</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>Both 443 (22)</td>
<td>5.4 (2.0)</td>
<td>505 (26)</td>
</tr>
<tr>
<td>Pitch</td>
<td>Neither</td>
<td>Neither 502 (30)</td>
<td>15.0 (2.2)</td>
<td>437 (29)</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Location 474 (29)</td>
<td>12.0 (3.2)</td>
<td>470 (30)</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td>Pitch 508 (30)</td>
<td>11.8 (3.1)</td>
<td>482 (30)</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>Both 426 (26)</td>
<td>8.2 (2.3)</td>
<td>513 (29)</td>
</tr>
<tr>
<td>Non-Spatial</td>
<td>Location</td>
<td>Neither 432 (32)</td>
<td>15.0 (3.0)</td>
<td>383 (22)</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Location 480 (28)</td>
<td>11.9 (2.7)</td>
<td>453 (23)</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td>Pitch 448 (29)</td>
<td>11.8 (2.2)</td>
<td>420 (24)</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>Both 396 (21)</td>
<td>8.2 (1.9)</td>
<td>410 (25)</td>
</tr>
<tr>
<td>Pitch</td>
<td>Neither</td>
<td>Neither 417 (29)</td>
<td>11.7 (2.2)</td>
<td>396 (28)</td>
</tr>
<tr>
<td></td>
<td>Location</td>
<td>Location 452 (28)</td>
<td>10.8 (3.1)</td>
<td>388 (29)</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td>Pitch 436 (29)</td>
<td>7.1 (3.0)</td>
<td>480 (29)</td>
</tr>
<tr>
<td></td>
<td>Both</td>
<td>Both 387 (25)</td>
<td>7.2 (2.2)</td>
<td>406 (28)</td>
</tr>
</tbody>
</table>
Table 2.3. Results of analysis of variance on mean reaction time of correct responses (RTs) and percentage of errors (PEs) for R2 in Experiment 2.

<table>
<thead>
<tr>
<th>Effect</th>
<th>RT</th>
<th>PE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>F</td>
</tr>
<tr>
<td>Response set (S)</td>
<td>456360.06</td>
<td>3.89</td>
</tr>
<tr>
<td>Task (T)</td>
<td>23002.41</td>
<td>0.96</td>
</tr>
<tr>
<td>Response (R)</td>
<td>5032.01</td>
<td>1.36</td>
</tr>
<tr>
<td>Pitch (P)</td>
<td>1022.49</td>
<td>0.49</td>
</tr>
<tr>
<td>Location (L)</td>
<td>56.20</td>
<td>0.02</td>
</tr>
<tr>
<td>T x R</td>
<td>816.12</td>
<td>0.17</td>
</tr>
<tr>
<td>T x P</td>
<td>28779.30</td>
<td>13.54***</td>
</tr>
<tr>
<td>T x L</td>
<td>33450.01</td>
<td>11.16**</td>
</tr>
<tr>
<td>R x P</td>
<td>82731.54</td>
<td>33.22***</td>
</tr>
<tr>
<td>T x R x P</td>
<td>7570.75</td>
<td>3.25</td>
</tr>
<tr>
<td>R x L</td>
<td>39527.31</td>
<td>15.10***</td>
</tr>
<tr>
<td>T x R x L</td>
<td>2687.12</td>
<td>1.39</td>
</tr>
<tr>
<td>P x L</td>
<td>145149.32</td>
<td>73.60***</td>
</tr>
<tr>
<td>T x P x L</td>
<td>10938.12</td>
<td>5.18*</td>
</tr>
<tr>
<td>R x P x L</td>
<td>5761.72</td>
<td>3.35</td>
</tr>
<tr>
<td>T x R x P x L</td>
<td>403.56</td>
<td>0.21</td>
</tr>
<tr>
<td>T x S</td>
<td>4988.43</td>
<td>0.21</td>
</tr>
<tr>
<td>R x S</td>
<td>6499.33</td>
<td>1.76</td>
</tr>
<tr>
<td>P x S</td>
<td>47.70</td>
<td>0.02</td>
</tr>
<tr>
<td>L x S</td>
<td>2038.26</td>
<td>0.75</td>
</tr>
<tr>
<td>T x R x S</td>
<td>4101.46</td>
<td>0.86</td>
</tr>
<tr>
<td>T x P x S</td>
<td>28.65</td>
<td>0.01</td>
</tr>
<tr>
<td>R x P x S</td>
<td>40.02</td>
<td>0.02</td>
</tr>
<tr>
<td>T x R x P x S</td>
<td>836.78</td>
<td>0.36</td>
</tr>
<tr>
<td>T x L x S</td>
<td>550.30</td>
<td>0.18</td>
</tr>
<tr>
<td>R x L x S</td>
<td>40724.50</td>
<td>15.55***</td>
</tr>
<tr>
<td>T x R x L x S</td>
<td>14871.09</td>
<td>7.70**</td>
</tr>
<tr>
<td>P x L x S</td>
<td>3782.71</td>
<td>1.92</td>
</tr>
<tr>
<td>T x P x L x S</td>
<td>4032.12</td>
<td>1.91</td>
</tr>
<tr>
<td>R x P x L x S</td>
<td>521.03</td>
<td>0.30</td>
</tr>
<tr>
<td>T x R x P x L x S</td>
<td>639.43</td>
<td>0.33</td>
</tr>
</tbody>
</table>

*df=(1,27), *p<.05, **p<.01, ***p<.001*
Let us consider the outcomes according to their theoretical implications. First we address the \textit{task effects} that reflect the impact of the task on the stimulus dimensions and the response. Second, we consider the \textit{stimulus-integration effects}; these effects are revealed by interactions between the stimulus features, showing that repetition of a particular feature enhances performance if the other feature is also repeated and hinders performance if the other feature is alternated. Third, we discuss \textit{stimulus-response-integration effects} by examining the interactions between repetition vs. alternation of the response and the stimulus features. Finally, we address \textit{response-set effects}.

\textit{Task effects.} There were two significant interactions in RT between task and location, and between task and pitch, showing that performance was facilitated in the location task by repeating a feature on the task-irrelevant dimension (439 msec vs. 470 msec respectively) or alternating the feature on the task-relevant dimension in the pitch task (441 msec vs. 471 msec respectively). In addition, the response interacted with the task in such a way that in the pitch task, responses were more accurate when being repeated than alternated (PEs 8.7\% vs. 11.2\% respectively) whereas, in the location task, alternation was more beneficial than repetition (PEs 8.9\% vs. 11.3\% respectively).

\textit{Stimulus-integration effects.} Pitch repetition interacted with location repetition, reflecting the standard crossover pattern with slower responses for trials in which one feature repeats while the other alternates, interestingly it was more prominent when the relevant feature repeated rather than alternated which may point to the role of attention in the process (see Figure 2.5). This interaction was also modified by task, suggesting that the pitch-location interaction was somewhat more pronounced in the location task than in the pitch task, but it was clearly reliable in both: \( F(1,27)=66.44, \ p<.0001, \) and \( F(1,27)=16.54, \ p<.0001, \) respectively.
This latter observation seems inconsistent with findings of Mondor and Leboe (2008), who failed to obtain interactions between pitch and location repetition when using a non-spatial response set. However, as pointed out earlier, they used a detection task that did not require the discrimination of any stimulus feature. This design choice was likely to prevent feature bindings from affecting performance in several ways. For one, it yielded average reaction times of less than 300 msec, which may have been too short to allow for the complete retrieval of the binding from the previous trial. Indeed, when Mondor and Leboe (2008) shortened the interval between prime and probe—a manipulation that they considered to facilitate binding retrieval and that effectively increased reaction times—a close-to-significant interaction between pitch- and location-repetition effects was obtained. Moreover, a detection task is likely to induce rather shallow perceptual coding processes, which again is likely to hamper the feature-matching process necessary to retrieve a particular binding. In any case, the present findings suggest that evidence for pitch-location binding can be obtained under favorable conditions.
To summarize, we were able to extend our observation of spontaneous pitch-loudness integration from Experiment 1 to the integration of pitch and location. Again, features from the two involved auditory dimensions were bound even though only one dimension was relevant at a time, suggesting that the mere temporal overlap of code activation is sufficient for integration.

**Stimulus-response-integration effects.** Analogously to Experiment 1, pitch and location repetition entered two-way interactions with response repetition, both in RTs and PEs, reflecting worse performance if a stimulus-feature repetition was accompanied by an alternation of the response, or vice versa (see Figure 2.6). The pitch-by-response interaction was unaffected by task, and a separate analysis confirmed that it was still reliable in the location task, both in RTs, $F(1,27)=9.04$, $p<.01$, and in PEs, $F(1,27)=12.70$, $p<.001$, as well as in the pitch task both in RTs $F(1,27)=27.10$, $p<.0001$, and in PEs, $F(1,27)=7.25$, $p<.05$. The location-by-response interaction was also unaffected by task. A reliable effect between location and response was observed in the pitch task both in RTs, $F(1,27)=4.3$, $p<.05$, and in PEs, $F(1,27)=8.614$, $p<.01$, as well as in location task both in RTs, $F(1,27)=15.41$, $p<.001$, and in PEs, $F(1,27)=8.56$, $p<.01$. 


Figure 2.6. Reaction times in Experiment 2 in the pitch task (upper panel) and the location task (lower panel) for relevant and irrelevant stimulus (repetition vs. alternation), as a function of response (repetition vs. alternation).

Response-set effects. The response set manipulation did not yield a reliable main effect in RTs nor PEs, even though participants tended to respond faster with a non-spatial than a spatial set—presumably reflecting the reduced
response uncertainty in the non-spatial Go/No-Go task. There were two reliable effects: the interaction between location and response was modified by response set (in RTs), and this three-way interaction was further modified by task (in RTs and PEs). Separate ANOVAs revealed that the location-response interaction was only reliable for the spatial response-set condition $F(1,13)=39.43, p<.0001$, rather than non-spatial response-set condition $F<1$, indicating stronger activation when the response include spatial features. Moreover, to disentangle the contributions to the four-way interaction, we analyzed the two tasks separately. In the location task, location and response repetition interacted significantly both in RTs, $F(1,27)=15.41, p<.001$; and in PEs, $F(1,27)=8.52, p<.01$, with no modulation by response set (see Figure 2.7, lower panel). However, in the pitch task, the location-response interaction was further modified by response set both in RTs, $F(1,27)=20.86, p<.0001$; and in PEs, $F(1,27)=9.94, p<.005$. Separate analyses of the pitch task by response set revealed significant location-response interactions only for the spatial response set both in RTs, $F(1,13)=19.16, p<.001$; and in PEs, $F(1,13)=24.47, p<.0001$ (see Figure 2.7, upper panel) and not in the non-spatial response set $F<1$ both in RTs and PEs. This pattern is in line with our expectation that a spatial response set amounts to making location task relevant, even with respect to stimulus coding. If location is task relevant by requiring that discrimination of S2 locations, location codes are strongly weighted anyway. As a consequence, stimulus location and responses are integrated, no matter whether the response set is spatially defined or not. However, when location is irrelevant with regard to S2 (in the pitch task, that is), location codes are weighted strongly only if location is relevant for discriminating the two responses but not if a non-spatial response set is used.
Figure 2.7. Reaction times in Experiment 2 in the pitch task (upper panel) and the location task (lower panel) for stimulus location (repetition vs. alternation), as a function of response (repetition vs. alternation) and response set group (spatial vs. non-spatial).
General Discussion

The aim of our study was to investigate the binding mechanism in and across auditory perception and action. In both experiments we found evidence for the spontaneous integration of auditory features: pitch and loudness in Experiment 1, and pitch and location in Experiment 2. Even though our participants were not instructed or required to create any feature conjunction, and even though nothing could be gained by doing so, the features of S1—a mere go signal—were apparently integrated into a coherent representation. This outcome is in line with previous findings in visual perception, where feature integration effects were obtained between shape and color, or shape and location (Hommel, 1998; Hommel & Colzato, 2004), and with findings from auditory studies, where evidence of integration was found for pitch and timbre (Hall et al., 2000; Takegata et al., 2005). We can conclude that feature-binding processes are not restricted to visual object perception; the modality targeted by FIT (Kahneman et al., 1992; Treisman & Gelade, 1980), but is following comparable principles in integrating auditory information.

Moreover, both experiments revealed interactions between stimulus and response indicative of stimulus-response feature binding. Again, these effects were obtained for all auditory dimensions investigated, that is, pitch, loudness, and location. These effects followed the same pattern as observed between stimulus features: repeating one member of a pair but not the other results in performance costs, usually in terms of RT and often in errors as well. This supports the idea that feature integration creates episodic links between the respective elements, which are retrieved as a whole when at least one element is encountered again (Hommel, 1998, 2004). This retrieval process does not take place if the relevant stimulus feature and the response are different from the previous ones, and it does not create any particular problem if all elements of the binding are repeated. In the case of partial repetitions (either the response or the
relevant stimulus feature), retrieval results in the reactivation of currently incorrect, conflicting information, and thus prolongs stimulus and/or response processing.

The fact that evidence for feature integration processes was obtained even under conditions where the processing of simple features would be sufficient supports the idea that integration occurs rather automatically. And yet, which information is integrated seems to be determined by the action goal. In particular, features that vary on dimensions that are relevant for defining a target stimulus or a response alternative are more likely to become part of bindings than features unrelated to such dimensions. Another principle underlying feature integration seems to be a temporal overlap of code activation. Codes of stimulus features seem to be processed sufficiently close in time to produce overlapping activations, even if the time needed to process them differs (see Experiment 1) and even if only one of them is task relevant. That is, features belonging to the same physical stimulus are likely to become part of the same object file. The integration of stimuli and responses is more sensitive to temporal characteristics. Stimuli that are closer in time to response execution seem to be more likely to be integrated with it. This fits with earlier observations of Hommel (2005), who found stimulus-response integration for stimuli presented briefly before, concurrently with, or even after the execution of the corresponding response, but no integration for stimuli presented during the planning of that response. Apparently, then, response execution provides the information necessary to trigger the integration process. A plausible candidate for pulling the trigger is the success of the response, which may signal that integrating the response with the apparently suitable context conditions is useful (Schultz, 2002). This possibility strengthened by the finding that integration of visual stimulus features and manual responses in a task like ours is facilitated by presenting positively-toned pictures after the execution of R1 (Colzato, van Wouwe, & Hommel, 2007a). Taken together, our findings provide evidence for the existence of temporary feature
binding in auditory perception and action, suggesting a general principle of how events are cognitively represented—presumably in terms of “event files”, as proposed by Hommel (1998, 2004).

Footnote

The temporal-overlap scenario sketched in Figure 2.4 refers to the hypothetical temporal relations between coding processes in Experiment 1. These relations depend on the particular stimuli and the stimulus parameters chosen and may thus look very different for other stimuli, intensities, and pitch values. We thus do not suggest that loudness is always coded faster than pitch or that loudness- and pitch-coding or pitch- and response-coding always overlap in time—we only suggest that features that happen to be coded by overlapping processes are more likely to be integrated.
Chapter 3

Intermodal Event Files: Integrating Features across Vision, Audition, Taction, and Action

Abstract

Understanding how the human brain integrates features of perceived events calls for the examination of binding processes within and across different modalities and domains. Recent studies of feature-repetition effects have demonstrated interactions between shape, color, and location in the visual modality and between pitch, loudness, and location in the auditory modality: repeating one feature is beneficial if other features are also repeated, but detrimental if not. These partial-repetition costs suggest that co-occurring features are spontaneously bound into temporary event files. Here, we investigated whether these observations can be extended to features from different sensory modalities, combining visual and auditory features in Experiment 1 and auditory and tactile features in Experiment 2. The same types of interactions, as for unimodal feature combinations, were obtained including interactions between stimulus and response features. However, the size of the interactions varied with the particular combination of features, suggesting that the salience of features and the temporal overlap between feature-code activations plays a mediating role.
Introduction

Human perception is multisensory, that is, we get to know our environment through multiple sensory modalities. The existence of multisensory perception raises the question of how the different sensory modalities’ features we process are integrated into coherent, unified representations. For example, eating an apple requires making sense of visual features such as the shape, color, and location of the fruit; a distinctive bite sound pattern of a particular pitch and loudness; a particular texture, weight, and temperature of the apple; and chemical features characterizing the apple’s taste and smell. These features are processed in distinct cortical regions and along different neural pathways (e.g., Goldstein, 2007), so that some mechanism is needed to bind them into a coherent perceptual representation—so as to solve what is known as the “binding problem” (Treisman, 1996). In the last decade, the investigation of binding processes has focused on visual perception (e.g., Allport, Tipper, & Chmiel, 1985; Treisman & Gelade, 1980) and only recently been extended to the auditory domain (e.g., Hall, Pastore, Acker, & Huang, 2000; Takegata et al., 2005). However, real objects are rarely defined in perceived in just one isolated modality but rather call for interactions among many sensory modalities. Therefore, an efficient feature-binding mechanism should operate in a multi-modal manner and bind features regardless of their modality.

In recent years, different research strategies were introduced to study multisensory perception. Some studies created situations of perceptual conflict such that two sensory modalities received incongruent information, which often produced perceptual illusions and, occasionally, even longer lasting aftereffects. A classic example is the McGurk effect in which vision changes speech perception: an auditory /ba/ sound is perceived as /da/ if paired with a visual lip movement saying /ga/ (McGurk & MacDonald, 1976). An additional audio-visual example is the ventriloquism effect: people mislocate sound sources after being
exposed to concurrent auditory and visual stimuli appearing at disparate locations (e.g., Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001). Another, more recently discovered illusion is the auditory-visual “double flash” effect in which a single visual flash is perceived as multiple flashes when accompanied by sequences of auditory beeps (Shams, Kamitani, & Shimojo, 2000). This illusion was also found in the auditory-tactile domain, where a single tactile stimulus leads to the perception of multiple tactile events if accompanied by tone sequences (Hötting & Röder, 2004). These and other studies in the multisensory domain provide evidence for on-line interactions between different sensory modalities, but they have not led to a comprehensive understanding of how the brain integrates those different features into coherent perceptual structures.

The purpose of the present study was to investigate multi-modal feature integration through the analysis of feature-repetition effects or, more precisely, of interactions between them. As Kahneman, Treisman, and Gibbs (1992), and many others since then, have shown, repeating a visual stimulus facilitates performance but more so if its location is also repeated. Further studies have demonstrated interactions between repetition effects for various visual and auditory features. For instance, repeating a visual shape improves performance if its color is also repeated but impairs performance if the color changes—and comparable interactions have been obtained for shape and location or color and location (Hommel, 1998; for an overview see Hommel, 2004). Auditory features interact in similar ways, as has been shown for sounds and locations (Leboe, Mondor, & Leboe, 2006) and pitch, loudness, and location (Zmigrod & Hommel, 2009).

The result patterns observed in these studies rules out an account in terms of mere priming. If repeating two features would simply produce better performance than repeating one feature or none, the most obvious interpretation would be that feature-specific priming effects are adding up to the best performance being associated with a complete repetition of the given stimulus.
Complete repetitions often yield comparable performance to “complete” alternations, that is, a condition where not a single feature repeats (e.g., Hommel, 1998). This implies that it is not so much that complete repetitions would be particularly beneficial but partial repetitions (repetitions of some but not all features of a stimulus) seem to impair performance. If we assume that co-occurring features are spontaneously integrated into an object file (Kahneman et al., 1992) or event file (Hommel, 1998), and that such files are automatically retrieved whenever at least some features of a stimulus are encountered again, we can attribute the observed partial-repetition costs to code conflict resulting from the automatic retrieval of previous but no longer valid features (Hommel, 2004). For instance, encountering a red circle after having processed a green circle may be difficult because repeating the shape leads to the retrieval of the just created <green+circle> binding, which brings into play the no longer valid color green. In any case, however, interactions between stimulus-feature-repetition effects are indicative of the spontaneous binding of features and thus can serve as a measure of integration.

**Aim of Study**

The main question addressed in the present study was whether comparable interactions can be demonstrated for combinations of features from different sensory modalities. We adopted the prime-probe task developed by Hommel (1998), which has been demonstrated to yield reliable integration-type effects for unimodal stimuli. It consists of trials (see Figure 3.1) in which two target stimuli are presented (S1 and S2) and two responses are carried out (R1 and R2). Most indicative of stimulus-feature integration is performance on R2, a binary-choice response to one of the features of S2, which is analyzed as a function of feature repetitions and alternations, that is, of the feature overlap between S1 (which commonly is more or less task irrelevant) and S2. Instead of unimodal stimuli we used binary combinations of visual and auditory stimuli (in Experiment 1) and of
auditory and vibro-tactile stimuli (in Experiment 2). The crucial question was whether the standard cross-over interaction patterns could be obtained with these multimodal feature combinations. If multimodal feature binding would occur just as spontaneously (as the present task does not require or benefit from integration) as in unimodal stimuli, we would expect that repeating a feature from one modality should improve performance if a feature from the other modality is also repeated, while performance should suffer if one feature is repeated but the other is not. In other words, we expected that partial repetitions would impair performance relative to complete repetitions or alternations.

A second question was whether task relevance has any impact on multimodal feature integration. From unimodal studies we know that task-relevant stimulus features are more likely involved in interaction effects. For example, if participants respond to the shape of S2 (while all features of S1 are entirely irrelevant and can be ignored), shape repetitions more strongly interact with other types of repetition, likewise color or location (e.g., Hommel, 1998). This suggests that making a feature dimension task-relevant induces some sort of top-down priming of that dimension, thus increasing the impact of repetitions on this dimension on the encoding and/or retrieval of feature bindings (Hommel, Memelink, Zmigrod, & Colzato, submitted). Our question was whether such task-relevance effects would also occur under multimodal conditions and we tested this question by manipulating task relevance within participants. Accordingly, they all served in two sessions, one in which one of the two features was task-relevant and one in which the other feature was relevant. We expected the repetition of the relevant feature would be more involved in interactions with other repetition effects indicative of feature integration.

A third question considered response repetition and its interactions with other repetition effects. Previous unimodal studies have revealed that stimulus features are apparently integrated with the response they accompany. For instance, having participants carry out a previously cued response (R1) to the
mere onset of the prime stimulus (S1), irrespective of any feature of that stimulus, induces similar interactions between repetition effects as observed between perceptual features. For instance, both repeating a stimulus feature and the response (e.g., if S1=S2 and R1=R2) and alternating the stimulus and the response yields far better performance than repeating the stimulus feature and alternating the response, or vice versa (e.g., Hommel, 1998). Again, the problem seems to be related to partial repetitions: repeating the stimulus feature or the response tends to retrieve the event file comprising of the previous stimulus-response combination, thus reactivating the currently no longer valid response or stimulus feature, respectively (Hommel, 2004). As comparable patterns have been obtained for both visual (e.g., Hommel, 1998) and auditory stimuli (e.g., Mondor, Hurlburt, & Thorne, 2003; Zmigrod & Hommel, 2009), we were interested to see whether they could also be obtained with multimodal stimuli. This was the reason why we complicated our design (which for stimulus-feature integration may do with S1, S2 and R2 alone) by having our participants carry out a prepared response (R1) to the mere onset of S1. Following Hommel (1998), we precued R1 in advance, so to ensure that S1 and R1 were entirely uncorrelated (so as to avoid associative learning or mapping effects). Nevertheless, we expected that the co-occurrence of S1 and R1 would suffice to create bindings between the features of S1 (in particular from the dimension that was relevant in S2) and R1, which should create interactions between the repetition effects of stimulus features and the response.
Figure 3.2. Sequence of events in Experiment 1, a visual response cue signaled a left or right mouse button click (R1) that was to be delayed until presentation of an audiovisual stimulus S1 (S1 is used as a detection signal for R1). The audiovisual stimulus S2 appeared 450 ms after R1. S2 signaled R2, a speeded left or right mouse button click according to the instructed mapping and task.

**Experiment 1**

Experiment 1 was performed to determine whether evidence for feature binding can be obtained for combinations of visual and auditory features and whether signs for stimulus-response binding can be obtained with multimodal stimuli. The visual stimuli and the tasks were adopted from Hommel’s (1998) design. The stimuli were combinations of a red or blue circle (color being the visual feature) and a pure tone of high or low pitch (the auditory feature). Participants were cued to prepare a response (left or right mouse button click), which they carried out (R1) to the onset of the first target stimulus (S1). The second stimulus (S2) appeared 450 ms after R1 response. Participants had to discriminate its color (in the color task) or pitch (in the pitch task) and carry out the response R2 (left or right mouse button click) assigned to the given feature value (see Figure 3.1).

We hypothesized that the pitch and color features of S1, although originating from different modalities, would still be bound when S2 was
encountered, so that any feature repetition would lead to the retrieval of that binding. This should create coding conflict with partial repetitions, so that impaired performance was expected for color repetitions combined with pitch alternations, and vice versa. Likewise, we expected that color and pitch (and the currently task-relevant feature in particular) would be integrated with the response, thus leading to interactions between color and response repetition and between pitch and response repetition.

One word of caution before going into the methodological details and the results: A major problem with multimodal stimuli, and often even with unimodal stimulus features, derives from the fact that different features are coded by different neural mechanisms, using different sensory transduction mechanisms and neural pathways, which leads to considerable and basically uncontrollable differences regarding processing speed and temporal dynamics (e.g., the time to reach a detection threshold and to decay), not to mention possible differences regarding salience and discriminability. As the temporal overlap between the coding of features seems to determine whether they interact (Hommel, 1993) and are integrated (Elsner & Hommel, 2001; Zmigrod & Hommel, 2009), the differences in temporal dynamics are likely to have consequences for the particular result patterns to be obtained. For instance, Hommel (2005) obtained evidence for stimulus-response integration only when stimuli appeared briefly before, simultaneously with, or even after the execution of the response, but not when stimuli appeared during the preparation of that response (i.e., when S1 accompanies the R1 cue). Along the same lines, Zmigrod and Hommel (2009) found more reliable effects of stimulus-response integration for stimuli that take longer to process and identify, so that they are coded closer in time to response execution. There is no obvious way to avoid the impact of temporal factors but they need to be taken into consideration in the interpretation of the results.
Method

Participants

Thirteen participants (2 men) recruited by advertisement served for pay or course credit. Their mean age was 21.5 years (range 18-28 years). All participants were naïve as to the purpose of the experiment and reported not having any known sight or hearing problems.

Apparatus and Stimuli

The experiment was controlled by a Targa Pentium 3, attached to a Targa TM 1769-A 17-inch CRT monitor. Participants faced the monitor at a distance of about 60 cm. The loudspeakers were located on both sides of the monitor at about 25 degrees left and right from the screen center, at a distance of about 70 cm to the participant. The bimodal target stimuli S1 and S2 were composed from two pure tones of 1000Hz and 3000Hz with duration of 50 ms and presented equally in both speakers at approximately 70 dB SPL, accompanied by a blue or red circle of about 10 cm in diameter. Responses to S1 and to S2 were made by clicking on the left or the right mouse button with index and middle fingers respectively. Response cues were presented in the middle of the screen (see Figure 3.1) with a right or left arrow indicating a left and right mouse click, respectively.

Procedure and Design

The experiment was composed of two sessions of about 20 min each. In the auditory session, pitch was the relevant feature and participants judged whether the pitch was high or low; in the visual session, color was the relevant feature and participants judged whether the color was blue or red. The order of sessions was counterbalanced across participants. Each session contained a practice block of 15 trials and an experimental block of 128 trials. The order of the trials was random. Participants were to carry out two responses per trial: the first response (R1) was a left or right mouse click to the onset of S1 (ignoring its
identity) as indicated by the direction of an arrow in the response cue, the second response (R2) was a left or right mouse click to the value of the relevant dimension of S2. Again, the identity of R1 was determined by the response cue and the time of execution by the onset of S1, whereas both identity and execution of R2 was determined by S2.

In the auditory session half of the participants responded to the high pitch (3000Hz) and the low pitch (1000Hz) by pressing on the left or right mouse button, respectively, while the other half received the opposite mapping. In the visual session half of the participants responded to the blue circle and to the red circle by pressing on the left or right mouse button, respectively, while the other half received the opposite mapping. The participants were instructed to respond as quickly and accurately as possible.

The sequence of events in each trial is shown in Figure 3.1. A response cue with a right or left arrow appeared for 1000 ms to signal R1, which was to be carried out as soon as S1 appeared. The duration between the response cue and S1 was 1000 ms. S2 came up 450 ms after R1, with the pitch (in the auditory session) or the color (in the visual session) signaling the second response (R2). In the case of incorrect or absent response an error message was presented on the screen. R2 speed and accuracy were analyzed as a function of session (visual vs. auditory), repetition vs. alternation of the response, and repetition vs. alternation of the visual feature (color), and repetition vs. alternation of the auditory feature (pitch).

Results

Trials with incorrect R1 responses (1%), as well as missing (RT>1200 ms) or anticipatory (RT<100 ms) R2 responses (0.9%) were excluded from analysis. The mean reaction time for corrected R1 was 290 ms (SD=87). From the remaining data, mean RTs and proportion of errors for R2 (see Table 3.1) were analyzed by means of four-way ANOVAs for repeated measures (see Table 3.2). We will present the outcomes according to their theoretical implications. First, we
address stimulus-repetition effects and interactions among them, which we consider evidence of *stimulus integration*. Second, we consider effects related to response repetition and interactions between response repetition and the repetition of stimulus features, which we assume to reflect *stimulus-response integration*.

**Table 3.1.** Experiment 1: Means of mean reaction time (RT in ms) and percentage of errors (PE) for R2, as a function of the relevant modality, the relationship between the stimuli (S1 and S2) and the relationship between the responses (R1 and R2).

<table>
<thead>
<tr>
<th>Attended Modality</th>
<th>The relationship between the stimuli (S1 and S2)</th>
<th>Response</th>
<th>Repeated</th>
<th>Alternated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RT</td>
<td>PE</td>
<td>RT</td>
</tr>
<tr>
<td>Visual</td>
<td>Color and Pitch Alternated</td>
<td>479</td>
<td>18.6</td>
<td>401</td>
</tr>
<tr>
<td></td>
<td>Only Color Repeated</td>
<td>425</td>
<td>6.6</td>
<td>446</td>
</tr>
<tr>
<td></td>
<td>Only Pitch Repeated</td>
<td>463</td>
<td>11.1</td>
<td>430</td>
</tr>
<tr>
<td></td>
<td>Color and Pitch Repeated</td>
<td>399</td>
<td>2.8</td>
<td>443</td>
</tr>
<tr>
<td>Auditory</td>
<td>Color and Pitch Alternated</td>
<td>518</td>
<td>18.1</td>
<td>428</td>
</tr>
<tr>
<td></td>
<td>Only Color Repeated</td>
<td>526</td>
<td>15.8</td>
<td>444</td>
</tr>
<tr>
<td></td>
<td>Only Pitch Repeated</td>
<td>457</td>
<td>6.4</td>
<td>516</td>
</tr>
<tr>
<td></td>
<td>Color and Pitch Repeated</td>
<td>430</td>
<td>3.1</td>
<td>494</td>
</tr>
</tbody>
</table>
Table 3.2. Experiment 1: Results of analysis of variance on mean reaction time (RT) of correct responses and percentage of errors (PE) of R2 in Experiment 1, df=(1,12) for all effects.

<table>
<thead>
<tr>
<th>Effect</th>
<th>RT</th>
<th>PE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>F</td>
</tr>
<tr>
<td>Task</td>
<td>87020.48</td>
<td>2.84</td>
</tr>
<tr>
<td>Response</td>
<td>7421.19</td>
<td>2.15</td>
</tr>
<tr>
<td>Pitch</td>
<td>776.48</td>
<td>0.46</td>
</tr>
<tr>
<td>Color</td>
<td>6000.87</td>
<td>3.53</td>
</tr>
<tr>
<td>Task * Response</td>
<td>8.10</td>
<td>0.01</td>
</tr>
<tr>
<td>Task * Pitch</td>
<td>6.39</td>
<td>0.00</td>
</tr>
<tr>
<td>Response * Pitch</td>
<td>107254.79</td>
<td>71.26 ***</td>
</tr>
<tr>
<td>Task * Response * Pitch</td>
<td>42242.13</td>
<td>13.60 **</td>
</tr>
<tr>
<td>Task * Color</td>
<td>907.23</td>
<td>0.33</td>
</tr>
<tr>
<td>Response * Color</td>
<td>29501.07</td>
<td>25.51 ***</td>
</tr>
<tr>
<td>Task * Response * Color</td>
<td>21564.50</td>
<td>20.60 ***</td>
</tr>
<tr>
<td>Pitch * Color</td>
<td>10522.23</td>
<td>8.89 **</td>
</tr>
<tr>
<td>Task * Pitch * Color</td>
<td>837.69</td>
<td>0.64</td>
</tr>
<tr>
<td>Response * Pitch * Color</td>
<td>532.61</td>
<td>0.15</td>
</tr>
<tr>
<td>Task * Response * Pitch * Color</td>
<td>261.86</td>
<td>0.37</td>
</tr>
</tbody>
</table>

* p<.05, **p<.01, ***p<.001

Stimulus integration. The RTs showed a significant interaction between color and pitch repetition. The effect followed the typical crossover pattern, with better performance for color repetition if pitch was also repeated than if it was alternated, but worse performance for color alternation if pitch was repeated than if it was alternated (see Figure 3.2). Separate ANOVAs, split by task, revealed that it was more pronounced in, and statistically restricted to the pitch task (pitch task: $F(1,12)=5.679$, $p<.05$; color task: $F(1,9)=2.796$, ns).
Figure 3.2. Reaction times of R2 in Experiment 1 as a function of repetition vs. alternation of the stimuli (S1-S2) of visual feature color and auditory feature pitch, regardless of the response.

Stimulus-response integration. The standard cross-over interactions between pitch and response repetition and between color and response repetition were found in RTs and error rates. As Figure 3.3 indicates, partial-repetition costs were obtained for both sensory modalities, that is, performance was impaired if a stimulus feature was repeated but not the response, or vice versa. These stimulus-response interactions were modified by task (i.e., the relevant modality), which called for more detailed analysis. Separate ANOVAs, split by task, revealed significant interactions between the stimulus feature from the relevant modality (i.e., pitch in the auditory task and color in the visual task) and the response in RT (visual task: $F(1,12)=43.11, p<.0001$; auditory task: $F(1,12)=45.97, p<.0001$) and errors (visual task: $F(1,12)=12.55, p<.005$; auditory task: $F(1,12)=32.24, p<.0001$). However, repeating the irrelevant stimulus (i.e., pitch in the visual task and color in the auditory task) interacted with response repetition only in the
visual task, thus producing a pitch-by-response interaction in RTs, $F(1,12)=4.89$, $p<.05$, and error rates, $F(1,12)=12.55$, $p<.005$; while no effects were obtained in the auditory task $F<1$.

Figure 3.3. Reaction times of R2 in Experiment 1 for repetition vs. alternation of the stimuli in the auditory-feature pitch and the visual-feature color, as a function of response repetition (vs. alternation) and task.
Discussion

Experiment 1 revealed interesting interactions between visual and auditory processes, and action planning. First, the findings demonstrate that performance depends on the repetition of combinations of visual and auditory features, suggesting an automatic integration mechanism binding features across attended and unattended modalities. This observation extends the findings from unimodal integration studies and supports the idea that feature integration is a general mechanism operating across perceptual domains.

Second, interactions between repetitions of stimulus features and responses were observed for both visual features (color) and auditory features (pitch). This replicates earlier findings from studies on visual coding and action planning (Hommel 1998, 2005) and on auditory coding and action planning (Mondor, Hurlburt, & Thorne, 2003; Zmigrod & Hommel, 2009), and supports the claim that binding mechanisms share codes across perception and action (Hommel, 1998).

Finally, consistent with previous observations from unimodal studies, we found that task relevance plays an important role in multimodal feature integration. At least stimulus-response integration was clearly influenced by which sensory modality was task-relevant, indicating that features falling on task-relevant dimensions are more likely to be integrated and/or retrieved. As suggested by Hommel (2004) and Zmigrod and Hommel (2009), task-relevant feature dimensions may be weighted more strongly (Found & Müller, 1996; Hommel, Müßeler, Aschersleben, & Prinz, 2001). Accordingly, the stimulus-induced activity of feature codes belonging to such a dimension will be stronger, thus increasing the amplitude of these codes and their lifetime (i.e., the duration they pass a hypothetical integration threshold). As a consequence, codes from task-relevant feature dimensions are more likely to reach the threshold for integration and to reach it for a longer time, which again makes them more likely
to be integrated with a temporally overlapping code and to overlap with a greater number of codes. This is particularly relevant for response-related codes, which reach their peak about one reaction time later than perceptual codes (assuming that response-code activation is locked to response onset the same way as stimulus-code activation is locked to stimulus onset). Only perceptual codes that are sufficiently strongly (and/or were sufficiently recently) activated, will survive this interval (Zmigrod & Hommel, 2009), which explains that task relevance is particularly important for stimulus-response integration.

In the present experiment, the temporal-overlap principal can account for stronger binding between task-relevant stimulus features and the response. It also may account for the observation that task-irrelevant pitch was apparently integrated with the response while task-irrelevant color was not. Given that in both tasks the responses were the same (mouse button click), the RT results show that participants were faster in the visual than the auditory task, suggesting that coding and identifying pitch took longer than coding and identifying color. Accordingly, pitch codes must have reached peak activation later than color codes. In the fast visual task, it means short time between the relatively late pitch-code activation and the response. While, in the slow auditory task, there is a long time between the relatively early color-code activation and the rather late response. Hence, the activation of the irrelevant pitch code was more likely to overlap with response activation than the activation of the irrelevant color code. It is true that at this point we are unable to rule out another possibility that is based on salience. As suggested by previous observations (Dutzi & Hommel, 2009), visual stimuli seem to rely much more on attention (and thus task-relevance) than auditory stimuli do—a phenomenon that has also been observed in other types of tasks (Posner, Nissen, & Klein, 1976). Hence, one may argue that auditory stimuli attract attention and are thus integrated irrespective of whether they are relevant for a task or not. However, Experiment 2 will provide evidence against this
posibility: even though auditory stimuli may well attract more attention, this does not necessarily mean that they are always integrated.

Experiment 2

Experiment 1 suggests that visual and auditory features are spontaneously bound both with each other and with the response they accompany, thereby extending similar observations from unimodal studies to multimodal integration. Experiment 2 was conducted to extend the range of features even further and to look into integration across audition, taction, and action. Even though experimental studies have often been severely biased towards vision, tactile perception plays an important role in everyday perception and interactions with our environment. Recent studies encourage the idea that tactile codes interact with codes from other modalities to create coherent perceptual states. For instance, vibrotactile amplitude and pitch frequency were found to interact in such a way that higher frequencies ‘feel’ more gentle (Sherrick, 1985; van Erp & Spapé, 2003). In the present study we used vibrotactile stimuli to create two different tactile sensations. This was achieved by using the Microsoft XBOX 360 controller, which produced either a ‘slow, rumbling’ vibration that was played by the pad’s low-frequency rotor, or a ‘fast, shrill’ one, by the pad’s high-frequency rotor. For the auditory feature we chose pitch, but to make sure that vibration rate did not interfere with perceiving acoustic frequencies, we used two tones of different shape (sinusoidal or square) but not period (1000 Hz), which were easily classified by participants as sounding either “clean” or “shrill”, respectively. The responses were also acquired by the Microsoft XBOX 360 controller.
Method

Participants
Ten participants (2 men) served for pay or course credit, their mean age was 20 years (range 18-27 years). All participants met the same criteria as in Experiment 1.

Apparatus and Stimuli
The same setup as in Experiment 1 was used, with the following exceptions. Instead of using the mouse we employed a Microsoft XBOX 360 gamepad which was connected to a Pentium-M based Dell laptop that communicated via serial port. The tactile features were based on two different rotors in the gamepad (low frequency vs. high frequency) for 500ms, and the auditory features were based on 1000Hz pitch with different shape (sinusoidal or square).

Procedure and Design
The procedure was as in Experiment 1, except for the following modifications. The visual task was replaced by the tactile task, in which participant had to judge whether the vibration rate is slow or fast. In addition, in the auditory task each participant had to judge whether the sound is clean or shrill. Moreover, the responses were acquired through the Microsoft XBOX 360 controller by having participants click with the right hand thumb on ‘A’ or ‘B’ buttons.
Results

The analysis followed the rationale of Experiment 1. Trials with incorrect R1 responses (0.5%), as well as missing (RT>1200 ms) or anticipatory (RT<100 ms) R2 responses (1.9%) were excluded from analysis. The mean reaction time for R1 was 219 ms (SD=91). Table 3.3 shows the means for RTs and proportion of errors obtained for R2. The outcomes of the ANOVAs for RTs and PEs are presented in Table 3.4.

Table 3.3. Experiment 2: Means of mean reaction time (RT in ms) and percentage of errors (PE) for R2 as a function of the relevant modality (auditory and tactile), the relationship between the stimuli (S1 and S2) and the relationship between the responses (R1 and R2).

<table>
<thead>
<tr>
<th>Attended Modality</th>
<th>The relationship between the stimuli (S1 and S2)</th>
<th>Response</th>
<th>Repeated</th>
<th>Alternated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditory</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pitch and Vibration Alternated</td>
<td>478</td>
<td>7.8</td>
<td>407</td>
</tr>
<tr>
<td></td>
<td>Only Pitch Repeated</td>
<td>483</td>
<td>6.6</td>
<td>425</td>
</tr>
<tr>
<td></td>
<td>Only Vibration Repeated</td>
<td>407</td>
<td>2.4</td>
<td>477</td>
</tr>
<tr>
<td></td>
<td>Pitch and Vibration Repeated</td>
<td>407</td>
<td>4.0</td>
<td>447</td>
</tr>
<tr>
<td>Tactile</td>
<td>Pitch and Vibration Alternated</td>
<td>608</td>
<td>19.8</td>
<td>551</td>
</tr>
<tr>
<td></td>
<td>Only Pitch Repeated</td>
<td>611</td>
<td>15.7</td>
<td>630</td>
</tr>
<tr>
<td></td>
<td>Only Vibration Repeated</td>
<td>639</td>
<td>15.4</td>
<td>604</td>
</tr>
<tr>
<td></td>
<td>Pitch and Vibration Repeated</td>
<td>503</td>
<td>9.8</td>
<td>568</td>
</tr>
</tbody>
</table>
### Table 3.4. Experiment 2: Results of analysis of variance on mean reaction time (RT) of correct responses and percentage of errors (PE) of R2 in df=(1,9) for all effects.

<table>
<thead>
<tr>
<th>Effect</th>
<th>RT MSE</th>
<th>RT F</th>
<th>PE MSE</th>
<th>PE F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>875895.10</td>
<td>12.93 **</td>
<td>1974.02</td>
<td>8.23 *</td>
</tr>
<tr>
<td>Response</td>
<td>437.55</td>
<td>.14</td>
<td>168.10</td>
<td>3.10</td>
</tr>
<tr>
<td>Pitch</td>
<td>12184.81</td>
<td>8.14 *</td>
<td>.62</td>
<td>.02</td>
</tr>
<tr>
<td>Vibration</td>
<td>5699.80</td>
<td>3.32</td>
<td>40.00</td>
<td>.84</td>
</tr>
<tr>
<td>Task * Response</td>
<td>117.63</td>
<td>.05</td>
<td>348.10</td>
<td>1.79</td>
</tr>
<tr>
<td>Task * Pitch</td>
<td>607.04</td>
<td>.37</td>
<td>18.22</td>
<td>.62</td>
</tr>
<tr>
<td>Response * Pitch</td>
<td>59354.31</td>
<td>12.41 **</td>
<td>792.10</td>
<td>.02 *</td>
</tr>
<tr>
<td>Task * Response * Pitch</td>
<td>18432.21</td>
<td>7.33 *</td>
<td>.40</td>
<td>.00</td>
</tr>
<tr>
<td>Task * Vibration</td>
<td>4232.38</td>
<td>1.33</td>
<td>10.00</td>
<td>.18</td>
</tr>
<tr>
<td>Response * Vibration</td>
<td>15759.51</td>
<td>5.79 *</td>
<td>70.22</td>
<td>.56</td>
</tr>
<tr>
<td>Task * Response * Vibration</td>
<td>23149.33</td>
<td>10.29 *</td>
<td>164.02</td>
<td>4.45</td>
</tr>
<tr>
<td>Pitch * Vibration</td>
<td>58549.66</td>
<td>32.38 ***</td>
<td>.90</td>
<td>.02</td>
</tr>
<tr>
<td>Task * Pitch * Vibration</td>
<td>25819.86</td>
<td>11.03 **</td>
<td>144.40</td>
<td>2.53</td>
</tr>
<tr>
<td>Response * Pitch * Vibration</td>
<td>219.70</td>
<td>.16</td>
<td>9.02</td>
<td>.40</td>
</tr>
<tr>
<td>Task * Response * Pitch * Vibration</td>
<td>2822.15</td>
<td>.82</td>
<td>27.22</td>
<td>.32</td>
</tr>
</tbody>
</table>

* p<.05, **p<.01, ***p<.001

First we will consider some effects of minor theoretical interest. A main effect of task in RTs and error rates was observed, indicating faster (441 vs. 589 ms) and more accurate (5.7% vs. 12.7%) performance in the auditory task. A main effect of pitch repetition was obtained, indicating faster responses with pitch repetitions than alternations (507 vs. 524 ms).

**Stimulus integration.** A significant interaction between pitch (repetition vs. alternation) and vibration rate (repetition vs. alternation) was obtained. This reflects a crossover pattern with slower responses for trials in which one feature repeats while the other alternates, as compared to complete repetitions or alternations (see Figure 3.4). This interaction was further modified by task, showing that it was more pronounced in, and statistically restricted to the
vibration task (vibration task: $F(1,9)=31.52, p<.001$; auditory task: $F(1,9)=2.09, ns.$).

![Graph showing reaction times of R2 in Experiment 2 as a function of repetition vs. alternation of the stimuli (S1-S2) of tactile feature vibration and auditory feature pitch, and task.]

**Figure 3.4.** Reaction times of R2 in Experiment 2 as a function of repetition vs. alternation of the stimuli (S1-S2) of tactile feature vibration and auditory feature pitch, and task.

*Stimulus-response integration.* There were significant interactions between pitch and response repetition as well as between vibration and response repetition in RTs. They followed the standard pattern of showing worse performance if the respective stimulus feature repeats while the response alternates, or vice versa. These two-way interactions were further modified by task (see Figure 3.5). Separate analysis revealed that the two-way interactions were reliable only for the task-relevant stimulus feature (response by pitch in the pitch task, $F(1,9)=17.14, p<.005$; response by vibration in the vibration task, $F(1,9)=26.51, p<.001$) but not for the task-irrelevant feature. In error rates, only the interaction between pitch and response repetition was reliable.
Figure 3.5. Reaction times of R2 in Experiment 2 for repetition vs. alternation of the stimuli in the auditory-feature pitch and the tactile-feature vibration, as a function of response repetition (vs. alternation) and task.

**Discussion**

Experiment 2 was successful in extending the evidence for visual-audio integration obtained in Experiment 1 to audio-tactile integration. Particularly clear was this evidence for the tactile task, where pitch and vibration were apparently bound automatically. Not so in the auditory task however. That may have to do with differences in salience, in the sense that the vibration stimulus was easier to ignore than the auditory stimulus. But it may also have to do with top-down
processes. Colzato, Raffone, and Hommel (2006) observed that the integration of stimulus features that differ in task relevance disappears with increasing practice, suggesting that participants learn to focus on the task-relevant feature dimension (and/or to gate out irrelevant feature dimensions). It may be that focusing on the auditory modality is easier or more efficient than focusing on the tactile modality, which may have worked against the integration of tactile information in the auditory task. In any case, however, we do have evidence that spontaneous audio-tactile integration can be demonstrated under suitable conditions.

Again, both features were integrated with the responses, only that now the task-relevance factor had an even more pronounced impact. Importantly, the observation that none of the task-irrelevant stimulus features was apparently bound with the response rules out the possibility that auditory stimuli always integrated—even if they may be more salient than others. This supports our interpretation that the asymmetries between modalities obtained in Experiment 1 reflect the temporal-overlap principle.

**General Discussion**

The aim of our study was to investigate whether features from different modalities are spontaneously bound both with each other and with the action they accompany. In particular, we asked whether cross-modality integration would be observed under conditions that in unimodal studies provide evidence for the creation of temporary object or event files. Experiment 1 provided evidence for the spontaneous integration across audition and vision and Experiment 2 for integration across audition and taction, suggesting that feature integration crosses borders between sensory modalities and the underlining neural structures. These findings fit with previous observations of interactions between sensory modalities, like in the McGurk effect or the flash illusion. However, they go beyond demonstrating mere on-line interactions in showing that the codes involved are bound into episodic multimodal representations that survive at least
half a second or so, as in the present study, and perhaps even longer (e.g., several seconds, as found in unimodal studies: Hommel & Colzato, 2004). One may speculate that these representations form the basis of multisensory learning and adaptation but supportive evidence is still missing. In the unimodal study of Colzato et al. (2006) participants were found to both learn and integrate combinations of visual features, but these two effects were independent. As pointed out by Colzato et al. and further developed by Hommel and Colzato (2009), this may suggest the existence of two independent feature-integration mechanisms: one being mediated by higher-order conjunction detectors or object representations; and the other by the ad-hoc synchronization of the neural assemblies coding for the different features. Along these lines, the present observations suggest that unimodal and multimodal ad-hoc binding operates in comparable ways.

A second aim or study was to investigate whether task relevance would play a similar role in multimodal integration as it does in unimodal integration. In particular, we expected that task-relevant features would be more likely to be involved in interactions with response features. This was in fact what we observed. Task relevance affected the binding between perceptual features and actions (in both experiments), and in some cases integration was actually confined to task-relevant stimuli and responses. Even though this observation strongly suggest that the handling of event files underlies considerable top-down control, the characteristics of our task does not allow us to disentangle two possible types of impact. On the one hand, the attentional set (reflecting the task instructions) may exclude irrelevant information from binding, suggesting that it is the creation of event files that is under top-down control. On the other hand, however, the effects we measure do not only require the creation of a binding but also its retrieval upon S2 processing, suggesting that control processes may operate on event-file retrieval. A recent study suggests that top-down control targets the retrieval rather than the creation of event files: If the task-relevance of features
changes from trial to trial, it is the attentional set assumed during S2 processing that determines the impact of a particular feature dimension but not the set assumed during S1 processing (Hommel, Memelink, Zmigrod, & Colzato, submitted). This suggests that the bindings that were created in the present study were comparable in the different tasks but the retrieval of previous bindings was (mainly) restricted to the features from task-relevant dimensions.

Apart from task relevance and attentional set, we found some evidence that the temporal dynamics of perceptual processing and, perhaps, the salience of stimuli affect the probability for a feature to be integrated and/or retrieved. In both experiments, the auditory feature was less dependent on task relevance than the features from other modalities. We considered two possible accounts, one in terms of temporal overlap and another in terms of salience. Given that both accounts are supported by other evidence, and given that the limited number of stimuli we used in our study does not allow us to disentangle the possible contributions, we do not consider these accounts as mutually exclusive and think that both temporal overlap and salience play a role that deserves further systematic investigation. Another possibly interesting observation is that, at least numerically, the cross modal visio-audio interaction was more pronounced in the auditory task and the cross modal audio-tactile interaction was more pronounced in the tactile task. In other words, the visual feature could not be ignored while attending the auditory feature and the auditory feature could not be disregarded when the task require attending to the tactile feature. Admittedly, this pattern of tactile > auditory > visual may merely reflect the particular dimensions and feature values that we picked for our study, but there is also another, theoretically more interesting possibility. Studies on the ontogenetic development of cortical multisensory integration show that the sensory modality-specific neurons in the midbrain mature in the very same chronological order (i.e., from tactile through audition to visual), which is also reflected in the sequence in which multisensory neurons emerge (Wallace, Carriere, Perrault, Vaughan, & Stein, 2006). It is thus
possible that the ontogenetic development of the sensory systems influence on the strength, the direction and the amount of connections among the sensory pathways.

Finally, we were interested to see whether multimodal stimuli would be integrated with the actions they accompany in the same way as unimodal stimuli are. Indeed, we replicated earlier findings suggesting audiomotor integration and extended that observation to the integration of tactile features with actions. As with other modalities, it was only particular features that interacted with the response but not whole stimulus events (which would have induced higher order interactions between both stimulus features and the response). As explained earlier, the possibility that task relevance affects retrieval only means that actions may very well be integrated with whole stimulus events but what is being retrieved is only the links between task-relevant elements. However, the possibility to do that suggests that bindings are not fully integrated structures that are activated in an all-or-none fashion but, rather, networks of links that are weighted according to task relevance (Hommel et al., 2001).

To sum up, our findings provide evidence for the existence of temporary feature binding across perceptual modalities and action, suggesting a rather general integration mechanism. Integration is mediated by task relevance, temporal overlap, and probably salience, but the same factors seem to be involved regardless of the modality or dimensions of the to-be-integrated features.
Chapter 4

Temporal Dynamics of Unimodal and Multimodal Feature Binding

Abstract

Two experiments studied the temporal dynamics of feature integration with auditory (Experiment 1) and audiovisual (Experiment 2) stimuli and manual responses. Consistent with previous observations, performance was better if the second of two consecutive stimuli shared all or none of the features of the first rather than if only one of the features overlapped; and comparable partial-overlap costs were obtained for combinations of stimulus features and the responses. These effects decreased systematically with increasing time between the two stimuli and response events, and the decrease rate was comparable for unimodal and multimodal bindings. General effect size reflected the degree of task relevance of the dimension or modality of the respective feature, but the effects of relevance and of temporal delay did not interact. This suggests that the processing of stimuli on task-relevant sensory modalities and feature dimensions is facilitated by task-specific attentional sets, whereas the temporal dynamics might reflect that bindings decay or become more difficult to access over time.
Chapter 4 – Temporal Dynamics

Introduction

One of the challenges human perception poses is to understand how the brain binds codes of features within and across sensory modalities, despite these being processed in various cortical areas (e.g., Goldstein, 2007; Wessinger et al., 2001; Zeki & Bartels, 1999). This so-called binding problem has been investigated initially in the visual domain (e.g., Allport, Tipper, & Chmiel, 1985; Kahneman, Treisman, & Gibbs, 1992; Treisman & Gelade, 1980), later in the auditory domain (e.g., Dyson & Quinlan, 2004; Hall, Pastore, Acker, & Huang, 2000; Takegata et al., 2005), and recently across modalities such as vision, audition, and taction (Zmigrod, Spapé, & Hommel, 2009). The available evidence suggests that binding mechanisms operate both within and across modalities and seem to bind perceptual features regardless of their origin.

Moreover, sequential-effects studies provide evidence that response-related features are also integrated with stimulus features into what Hommel (1998, 2004) has called ‘event files’, that is integrated episodic traces of all the perceptual and action features related to a particular event. In these studies, participants typically carry out two responses in a row (see Figure 4.1). First they see a response cue that signals the first response (R1), which however is to be carried out only after a trigger stimulus (S1) is presented. After a short SOA (stimulus-onset asynchrony) or RSI (response-stimulus interval) the second stimulus (S2) appears and calls for a binary-choice response to one of its features (R2). Similar to the findings from visual and auditory studies, main effects of stimulus-feature repetition were obtained. Yet more interesting, interactions between different stimulus-feature-repetition effects and between stimulus- and response-repetition effects were observed for visual features and response (Hommel, 1998, 2005), auditory features and response (Mondor, Hurlburt, & Thorne, 2003; Zmigrod & Hommel, 2009), and tactile features and response (Zmigrod et al., 2009).
These observations suggest that stimulus and response features are spontaneously integrated into multimodal event files, which are retrieved whenever at least one feature repeats. Assume, for instance, S1 and S2 consist of varying combinations of auditory pitch (low vs. high) and visual color (red vs. blue), as in the study of Zmigrod et al. (2009). Findings show that a complete repetition of both features (S1=blue+low → S2=blue+low, say) or a “complete” alternation (red+high → blue+low) produce better performance than partial repetitions (red+low → blue+low or blue+high → blue+low). This suggests that the combination presented as S1 is automatically integrated and retrieved upon repetition of any feature. If no retrieval occurs (as with complete alternations) or the retrieved episode fits with the present combination (as with complete repetitions), the processing of the current episode is unimpaired. In case of a partial repetition, however, retrieval reactivates a code that now competes with coding the present stimulus (e.g., if S1=red+low and S2=blue+low, the repetition of the “low” feature reactivates the S1 episode, which brings into play the “red” that now competes with coding for blue). In other words, partial repetitions induce code conflict (Hommel, 1998, 2004).

Empirical evidence suggests that attention is not strictly required for binding. For instance, Mordkoff and Halterman (2008) showed with a modified flanker task—where to-be-ignore flankers were correlated with responses—that stimulus-response bindings can be created without attention. In the auditory domain, Takegata and colleagues (2005) have shown that feature integration can occur without focused attention on the to-be-integrated features. In their EEG study, participants performed a visual working memory task while ignoring a background of two sounds that varied in timbre and pitch. The pitch-timbre combinations elicited similar amplitudes and latencies in the ERP component mismatch negativity. Likewise, Dyson & Quinlan (2004) reported that responses to consecutive auditory stimuli composed of pitch-location combinations were slower when there was a change in the irrelevant dimension. In line with this
view, studies demonstrated that irrelevant features from the same modality as the attended feature (in vision: Hommel, 2005; in audition: Dyson & Quinlan, 2004), or from a non-attended modality (Zmigrod et al., 2009) can be bound to a response. This suggests that binding can occur implicitly and automatically, even though other findings suggest that bindings involving attended features, or features from attended dimensions, affect performance more strongly and reliably than bindings involving unattended features (e.g., Hommel, 2005; Hommel & Colzato, 2009; Talsma & Woldorff, 2005). The creation and/or the retrieval of bindings thus seems to benefit from, but does not require focused attention.

Multisensory perception has received increasing attention in the last two decades or so and many studies have investigated the interplay between different modalities, mostly in conflict situations. These, in turn, produce illusions such as the McGurk effect (McGurk & MacDonald, 1976), spatial or temporal ventriloquism (e.g., Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen, Bertelson, & de Gelder, 2001; Vroomen & de Galder, 2004), or the double flash effect (Hötting & Röder, 2004; Shams, Kamitani, & Shimojo, 2000), but how intermodal binding actually works is still unclear. Multimodal perception (such as with audio-visual stimuli) faces ‘binding problems’ that are far more complicated than within a single modality, due to the fundamental differences both in the physical properties of, say, sound and light and in the sensory transduction mechanisms (e.g., in transduction latencies which prevents the use of tight temporal-synchrony criteria for cross-modal binding). And yet, our conscious perception of multimodal stimuli is commonly coherent and unified, suggesting that binding works. Different properties were tested as to provide the glue which binds features from different modalities together, such as physical and conceptual relationships between successive events (Gordon & Irwin, 2000), cohesion (Mitroff, Scholl, & Wynn, 2004), context (Mitroff, Arita, & Fleck, 2009), and spatiotemporal continuity (Gao & Scholl, 2010; Mitroff, & Alvarez 2007). Yet,
no conclusive conclusion can be drawn at this point, perhaps because there is not just one criterion but many (Massaro, 1987).

Hence, even though there is evidence suggesting that spontaneous binding takes place in and across various perceptual and action modalities, it is still unclear exactly how bindings are created and retrieved; and whether the respective processes are the same within modality and across modalities. In the present study, we compared the temporal dynamics of unimodal and multimodal bindings in order to identify commonalities—which would suggest a common mechanism—and differences—which would point to separable mechanisms. We used the same task as in previous studies on unimodal and multimodal integration (e.g., Zmigrod, et al., 2009) but extensively varied the RSI between the event that is supposed to induce binding (S1 and R1) and the event that leads to the later retrieval of this binding (S2 and R2). (Obviously, bindings may also be retrieved by the former and be created by the latter, but our design was balancing these effects out.) This manipulation was thought to tap into the robustness and stability of the bindings created upon processing the episode comprising of S1 and R1 and/or the accessibility of these bindings for retrieval. Previous investigations of the impact of the stimulus interval on the aftereffects of unimodal visual bindings (Hommel & Colzato, 2004) have provided evidence for both relative stability (i.e., the relationship between the two events matters even with intervals of 4 seconds) and decay (FOOTNOTE 1) (i.e., effect sizes tended to decrease over time). Considering these observations our present study of unimodal auditory binding (Experiment 1) and multimodal audiovisual binding (Experiment 2) was aimed at addressing three questions of theoretical relevance:

First, we were interested to see whether there are comparable binding effects between the auditory domain and the visual domain over time. To investigate that, we compared the decay rates of bindings involving auditory stimulus features with those involving visual stimulus features, by making loudness the relevant stimulus feature in Experiment 1 and color the relevant
stimulus feature in Experiment 2, and comparing the effects of the two modalities under similar conditions.

Second, we wished to explore the role of attention both in unimodal and multimodal feature binding over time. To investigate that, we examined whether the decay rates of bindings involving relevant and irrelevant features would be comparable. Previous findings with unimodal visual stimuli have shown that task relevance is an important factor for how sizable and reliable feature bindings are. For instance, making the shape task relevant by mapping R2 (which is a binary choice response) to the shape of S2 has been found to induce strong interactions between shape repetition and response repetition (the statistical indicator of the binding process) but only weak interactions between color repetition and response repetition (Hommel, 1998). Changing the instruction by mapping R2 onto the color of S2 reversed this pattern and led to weak interactions between shape and response repetition and strong interactions between color and response repetition. This suggests that making a feature relevant leads to the stronger weighting of the respective feature dimension, which increases the impact of features falling on this dimension on performance (Hommel & Colzato, 2009). This may affect decay of bindings involving relevant and irrelevant features in two different ways: bindings with relevant features may decay less (i.e., the decay rate may differ), or they may show the same decay rate as bindings with irrelevant features but start decaying from an initially higher level. In other words, the task relevance of the features involved may affect either the slope or the intercept of the RSI function of binding effects.

Third, we were interested to see whether attention would affect unimodal and multimodal bindings alike. Our two experiments were designed in such a way that they differed with respect to the task-relevant stimulus feature and its sensory modality (auditory in Experiment 1 and visual in Experiment 2) but were comparable with respect to the task-irrelevant stimulus feature, which was auditory pitch in both cases. This allowed for testing whether modality-specific
attentional sets, which we assumed to differ between the two experiments, would affect the binding between the irrelevant pitch with the relevant stimulus feature and with the response (which was the same in both experiments). Given that audition was task relevant in Experiment 1, which presumably led to the allocation of more attentional resources to the auditory modality, the processing of pitch may benefit from that and thus allow for a more effective creation and/or retrieval of pitch-related bindings.

The Task

In this study, we used the event file paradigm (Hommel, 1998) both in the unimodal experiment with auditory stimuli (Experiment 1) and the multimodal experiment with audio-visual stimuli (Experiment 2). To examine the temporal dynamics of bindings, we varied the temporal distance between the creation of bindings (S1/R1) and their assumed automatic retrieval (induced by S2/R2 processing). As depicted in Figure 4.1, each trial started with the presentation of a response cue in the form of a directional arrow, indicating whether a left or right response was required to the mere onset of S1 (regardless of its features). S2 appeared either 500, 1000, 2000, 3000, or 4000 ms after responding to S1, depending on the RSI (which varied between participants). The response to S2 (R2) was a binary choice reaction according to the task relevance feature of S2 (loudness in Experiment 1, loud vs. soft; and color in Experiment 2, red vs. blue).

In each experiment, two stimulus features and one response feature varied independently, so that feature-repetition effects (reflecting the relationship between S1 and S2 and between R1 and R2) and, more importantly, their interactions (taken to indicate bindings) could be studied. One of the stimulus features was task-relevant by signaling R2 (loudness in Experiment 1 and color in Experiment 2) and one was task-irrelevant (pitch in both experiments), while the response was always task-irrelevant. Feature-repetition effects (calculated by performance if the feature alternated minus performance if the feature was
repeated) could interact in three ways (apart from a possible three-way interaction): the repetition effects for the relevant and irrelevant stimulus features could interact, which would signal perceptual binding (Treisman & Gelade, 1980); and each of the two stimulus-repetition effects could interact with the response-repetition effect, which would signal stimulus-response binding (Hommel, 1998, 2004).

**Experiment 1**

Experiment 1 used unimodal stimuli and manual responses, with loudness being the relevant and pitch the irrelevant stimulus feature. The interval between the response to the first event and the onset of the second (RSI) was varied between 500 and 4000 ms (amounting to SOAs of about 800-4300 ms), similar to the unimodal visual study of Hommel and Colzato (2004).

**Figure 4.1.** Sequence of events in Experiment 1 (upper panel) and Experiment 2 (lower panel). A visual response cue signaled a left or right mouse button click (R1) that was to be delayed until presentation of the first stimulus S1 (S1 is used as a detection signal for R1). The second stimulus S2 appeared either: 500, 1000, 2000, 3000, or 4000 ms after R1 depending on the group. S2 signaled R2, a speeded left or right mouse button click according to the instructed mapping and task.
Method

Participants

Fifty-five participants (7 men), students at Leiden University, were recruited by advertisement and were paid or received a course credit for a 20 min session. Their mean age was 20 years (range 18-27 years). All the participants reported not having any known sight or hearing problems. The participants were naïve as to the purpose of the experiment. The participants were randomly but equally assigned to five groups with different RSIs (500, 1000, 2000, 3000, and 4000 ms).

Apparatus and Stimuli

The experiment was controlled by a Targa Pentium 3, attached to a Targa TM 1769-A 17-inch monitor. Participants faced the monitor at a distance of about 60 cm along with headphones. The experiment’s stimuli S1 and S2 were composed from two pure tones of 1000Hz and 3000Hz with duration of 50 ms, and were presented at 60 and 80 dB SPL. Response cues were presented in the middle of the screen (see Figure 4.1) with right or left arrows indicating a left and right mouse click, respectively. Responses to S1 (serving as mere go signal) and to (the loudness of) S2 were made by clicking on the left or the right mouse button with the same hand.

Procedure and Design

The experiment was composed of a practice block with 15 practice trials and an experimental block with 96 experimental trials. The order of the trials was randomized. Participants had to carry out two responses per trial: R1 was a simple reaction with a left or right mouse click as indicated by the direction of an arrow in the response cue. It had to be carried out as soon as S1 appeared, regardless of its loudness or pitch. R2 was a binary-choice reaction to the loudness of S2. Half of the participants responded to the loud and soft sound by pressing on the left or
right mouse button, respectively, while the other half received the opposite mapping. The participants were guided to respond as quickly and accurately as possible.

The sequence of events in each trial is shown in Figure 4.1 (upper panel). A response cue with a right or left arrow was visually presented for 1500 ms signaling response (R1) which was to be carried out after stimulus 1 was presented. S2 appeared either: 500, 1000, 2000, 3000, or 4000 ms after the onset of R1, the response to S1. In case of incorrect or absent response an error message was presented.

**Results and Discussion**

Trials with incorrect R1 responses (0.3%), as well as missing (RT>1200 ms) or anticipatory (RT<100 ms) R2 responses (0.05%) were excluded from analysis. The mean reaction time for correct R1 was 284 ms (SD=80). From the remaining data, mean reaction time (RTs) and percentage of errors (PEs) for R2 were analyzed as a function of the four variables: the relationship between S1 and S2 (repetition vs. alternation) with regard to loudness and pitch, the relationship between responses R1 and R2 (repetition vs. alternation); and the RSI condition (500, 1000, 2000, 3000, or 4000 ms). ANOVAs were performed by using a mixed design with repeated measures on three variables and with RSI as between group variable.

The analysis revealed significant main effects of the repetition of loudness, the relevant feature, in RTs, $F(1,50)=13.47$, $p<.001$; and PEs, $F(1,50)=21.17$, $p<.001$, indicating slower responses and more errors in repeating trials (553 ms and 9.4% respectively) than alternating trials (532 ms and 5% respectively). Additionally, there was a significant main effect of the repetition of pitch, the irrelevant feature, in RTs, $F(1,50)=12.61$, $p<.001$, indicating a quicker response in repetition trials (532 ms) than alternation trials (552 ms). Replicating earlier findings (see Zmigrod & Hommel, 2009), the results further revealed
interactions between the relevant and the irrelevant stimulus feature and between stimulus and response features, which can be taken to reflect the aftereffects of stimulus-feature binding and stimulus-response binding, respectively (Hommel, 1998). In the following, we group our observations according to these theoretical implications.

First, there were a number of effects involving the repetition of loudness and pitch, the relevant and the irrelevant stimulus feature (for means see Table 4.1). A significant interaction between pitch repetition and loudness repetition in RTs, $F(1,50)=44.27$, $p<.001$, indicated that repeating one auditory feature impairs performance if the other auditory feature is alternated rather than repeated—the standard observation in sequential-effect studies (e.g., Hommel, 1998). This interaction was further modified by response repetition in RTs, $F(1,50)=9.87$, $p<.005$; and PEs, $F(1,50)=29.44$, $p<.001$. These three-way interactions were due to particularly good performance if all three features were either repeated or alternated—a common pattern that has been attributed to shortcutting response selection processes with complete repetitions (Bertelson, 1963) and alternations (Hommel & Colzato, 2004). In case of the PEs, the interaction was further modified by RSI, $F(4,50)=3.29$, $p<.05$, reflecting that this shortcutting pattern was more pronounced in the short RSIs.

Table 4.1. Experiment 1: Means of mean reaction times for responses to stimulus 2 (RT$_{R2}$ in ms) as a function of the response-stimulus interval (RSI in ms), and the relationship between the stimuli features (S1 – S2) for loudness and pitch.

<table>
<thead>
<tr>
<th>RSI (ms)</th>
<th>Loudness repeated</th>
<th>Loudness alternated</th>
<th>Partial repetition cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pitch repeated</td>
<td>Pitch alternated</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>565</td>
<td>645</td>
<td>49</td>
</tr>
<tr>
<td>1000</td>
<td>516</td>
<td>563</td>
<td>40</td>
</tr>
<tr>
<td>2000</td>
<td>476</td>
<td>538</td>
<td>37</td>
</tr>
<tr>
<td>3000</td>
<td>516</td>
<td>577</td>
<td>28</td>
</tr>
<tr>
<td>4000</td>
<td>554</td>
<td>578</td>
<td>20</td>
</tr>
</tbody>
</table>
Second, the effects of loudness (the relevant stimulus feature) and response repetition (for means see Table 4.2) interacted in RTs, $F(1,50)=60.91$, $p<.001$; and in PEs, $F(1,50)=21.25$, $p<.001$. Performance was impaired if loudness was repeated but the response alternated, or vice versa, thus replicating earlier observations (Zmigrod & Hommel, 2009).

Third, the effects of pitch (the irrelevant stimulus feature) and response repetition (for means see Table 4.2) interacted in RTs, $F(1,50)=23.66$, $p<.001$; and PEs, $F(1,50)=39.81$, $p<.001$, due to worse performance if pitch was repeated while the response alternated, or vice versa. The interaction in PEs was further modified by RSI, $F(4,50)=3.71$, $p<.01$. As revealed by separate analyses, pitch and response repetition interacted reliably in all but the longest RSI ($Fs(1,10)=26.26$, $p<.001$; 6.74, $p<.05$; 5.43, $p<.05$; 13.5, $p<.005$; $F<1$; respectively).

Table 4.2. Experiment 1: Means of mean reaction times for responses to stimulus 2 (RT\textsubscript{R2} in ms) as a function of the response-stimulus interval (RSI in ms), the relationship between the responses (R1 and R2), and the relationship between the stimuli features (S1 and S2) for loudness and pitch.

<table>
<thead>
<tr>
<th>RSI (ms)</th>
<th>Response repeated</th>
<th>Response alternated</th>
<th>Partial repetition cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Loudness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>repeated</td>
<td>alternated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Loudness</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>repeated</td>
<td>alternated</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>560</td>
<td>597</td>
<td>650</td>
</tr>
<tr>
<td>1000</td>
<td>514</td>
<td>552</td>
<td>565</td>
</tr>
<tr>
<td>2000</td>
<td>489</td>
<td>484</td>
<td>525</td>
</tr>
<tr>
<td>3000</td>
<td>531</td>
<td>543</td>
<td>563</td>
</tr>
<tr>
<td>4000</td>
<td>552</td>
<td>556</td>
<td>581</td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>repeated</td>
<td>alternated</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pitch</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>repeated</td>
<td>alternated</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>542</td>
<td>616</td>
<td>613</td>
</tr>
<tr>
<td>1000</td>
<td>510</td>
<td>556</td>
<td>554</td>
</tr>
<tr>
<td>2000</td>
<td>462</td>
<td>511</td>
<td>493</td>
</tr>
<tr>
<td>3000</td>
<td>511</td>
<td>563</td>
<td>532</td>
</tr>
<tr>
<td>4000</td>
<td>545</td>
<td>563</td>
<td>564</td>
</tr>
</tbody>
</table>
Taken together, the effects we obtained support previous findings regarding feature integration in and across perception and action planning in general (Hommel, 2004) and regarding auditory perception and manual action planning in particular (Zmigrod & Hommel, 2009). However, the main focus of the present study is on the temporal dynamics of integration effects, and the processes underlying them, and thus on the changes of integration effects over time. In order to analyze and directly compare these temporal patterns, we took the interaction terms to calculate partial repetition costs ($\text{RT}_{\text{partial repetition}} - \text{RT}_{\text{complete repetition/alternation}}$) per effect type (FOOTNOTE 2) and RSI, and ran an ANOVA with RSI and effect type (loudness-pitch, loudness-response, and pitch-response) as factors. The only result this produced was a significant main effect of RSI, $F(4,50)=3.01, p<.05$. We take these observations to suggest that the costs produced by partial repetition decrease over time for all types of effects, and they do so in rather comparable ways (Figure 4.2). In addition, comparing the effect types through an ANOVA with RSI and effect type as factors revealed a main effect of effect type that approached significance, $F(1,50)=3.74, p=.059$. This not-quite-reliable trend reflects more pronounced partial-repetition costs if the relevant stimulus feature is involved, which is consistent with previous hints towards a role of task relevance in boosting the impact of bindings on performance (Hommel, 1998). If anything, however, this role does not seem to be strong with the present auditory stimuli. With regard to the temporal dynamics, our findings are in line with observations in the visual domain (Hommel & Colzato, 2004), which also indicated that bindings affect performance within a temporal window of at least 3–4 sec.
Experiment 1 provided evidence that bindings involving auditory features are relatively robust, but there was also substantial decay over time. Experiment 2 went on to see whether a comparable pattern could be found for intermodal integration. Accordingly, we replaced the auditory feature loudness by the visual feature color as the task-relevant variable, that is, as the variable that signaled R2.

**Method**

**Participants**

Fifty-five participants, students at Leiden University (5 men) with a mean age of 20.5 years (range 18-30 years) participated for a 20 min session. All the participants reported not having any known sight or hearing problems. The participants were naïve as to the purpose of the experiment. The participants were
randomly assigned to five groups with different RSIs (500, 1000, 2000, 3000, and 4000 ms).

**Apparatus and Stimuli.**

The apparatus was as in Experiment 1. The auditory stimuli were pure tones of 1000Hz and 3000Hz with duration of 50 ms presented at approximately 70 dB SPL, and they were accompanied by a blue or red circle in the middle of the monitor.

**Procedure and Design**

The same setup as in Experiment 1 was used with the following exceptions. The task stimuli were audiovisual, and the task for R2 was a binary-choice reaction to the color of S2. The experiment was composed of a practice block with 15 practice trials and an experimental block with 128 experimental trials. The sequence of events in each trial is shown in Figure 4.1 (lower panel).

**Results and Discussion**

Trials with incorrect R1 responses (1%), as well as missing (RT>1200 ms) or anticipatory (RT<100 ms) R2 responses (0.05%) were excluded from analysis. The mean RT for correct R1 was 239 ms ($SD=75$). Analogously to Experiment 1, mean RTs and PEs for R2 were analyzed as a function of the four variables: the relationship between S1 and S2 (repetition vs. alternation) with regard to color and pitch, the relationship between responses R1 and R2 (repetition vs. alternation); and the RSI condition (500, 1000, 2000, 3000, or 4000 ms).

Similar to Experiment 1, there was a significant main effect of pitch in RTs, $F(1,50)=5.38, p<.05$, due to faster responses when pitch was repeated (466 ms) than when it was alternated (475 ms). Additionally, a significant main effect of response in PEs was obtained, $F(1,50)=4.88, p<.05$, due to more accurate responses on alternation trials (7.6% ) than on repetition trials (8.9%). But, again, the interaction effects were of greater theoretical interest.
First, we analyzed the effects involving color and pitch, the relevant and the irrelevant stimulus feature (for means see Table 4.3). There was not reliable interaction between color and pitch in the overall analysis, and the three-way interaction involving RSI also failed to reach significance. However, separate analyses for the five RSIs revealed a significant interaction of color and pitch repetition in RTs in the shortest interval (500 ms), $F(1,10)=6.51, p<.05$. This result conceptually replicates the finding of Zmigrod et al. (2009), who also used a rather short interval between the events. However, in comparison to Experiment 1, it is notable that the effect of the multimodal color-pitch binding is much less pronounced than that of the unimodal loudness-pitch binding.

<table>
<thead>
<tr>
<th>RSI (ms)</th>
<th>Color repeated</th>
<th>Color alternated</th>
<th>Partial repetition cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pitch repeated</td>
<td>Pitch alternated</td>
<td>Pitch repeated</td>
</tr>
<tr>
<td>500</td>
<td>438</td>
<td>456</td>
<td>465</td>
</tr>
<tr>
<td>1000</td>
<td>467</td>
<td>478</td>
<td>457</td>
</tr>
<tr>
<td>2000</td>
<td>421</td>
<td>441</td>
<td>410</td>
</tr>
<tr>
<td>3000</td>
<td>514</td>
<td>519</td>
<td>511</td>
</tr>
<tr>
<td>4000</td>
<td>494</td>
<td>495</td>
<td>468</td>
</tr>
</tbody>
</table>

Second, the effects of color and response repetition (for means see Table 4.4) interacted in RTs, $F(1,50)=67.04, p<.001$; and in PEs, $F(1,50)=49.32, p<.001$. Performance was impaired if color was repeated but the response alternated, or vice versa, consistent with earlier observations (Hommel, 1998). The PE interaction was further modified by RSI, $F(4,50)=4.41, p<.01$. Separate analyses revealed significant color-response interactions for all but the longest RSIs ($F$s$(1,10)=10.54, p<.01; 23.00, p<.001; 6.52, p<.05; 18.71, p<.01; F<1$, respectively).
Third, the effects of pitch and response repetition (for means see Table 4.4) interacted in RTs, $F(1,50)=4.51, p<.05$; and in PEs, $F(1,50)=19.26, p<.001$, indicating worse performance with partial repetitions of either pitch or the response as compared to complete repetitions or alternations. The PE interaction was further modified by RSI, $F(4,50)=2.83, p<.05$. Separate analyses revealed significant interactions in the 500 and 1000 ms RSIs ($F(1,10)=10.25, p<.01$; $F(1,10)=19.39, p<.001$, respectively), but not in the other RSI conditions. It is notable that the pitch-response interactions were considerably less pronounced than in Experiment 1, indicating that pitch-response bindings were less robust and more transient. We will get back to this issue.

**Table 4.4.** Experiment 2: Means of mean reaction times for responses to stimulus 2 (RT$_{R2}$ in ms) as a function of the response-stimulus interval (RSI in ms), the relationship between the responses (R1 and R2), and the relationship between the stimuli features (S1 and S2) for color and pitch.

<table>
<thead>
<tr>
<th>RSI (ms)</th>
<th>Color repeated</th>
<th>Color alternated</th>
<th>Color repeated</th>
<th>Color alternated</th>
<th>Partial repetition cost</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Response repeated</td>
<td>Response alternated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>425</td>
<td>495</td>
<td>469</td>
<td>428</td>
<td>56</td>
</tr>
<tr>
<td>1000</td>
<td>445</td>
<td>475</td>
<td>501</td>
<td>440</td>
<td>45</td>
</tr>
<tr>
<td>2000</td>
<td>416</td>
<td>429</td>
<td>446</td>
<td>401</td>
<td>29</td>
</tr>
<tr>
<td>3000</td>
<td>501</td>
<td>535</td>
<td>532</td>
<td>498</td>
<td>34</td>
</tr>
<tr>
<td>4000</td>
<td>480</td>
<td>496</td>
<td>509</td>
<td>456</td>
<td>35</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Pitch repeated</th>
<th>Pitch alternated</th>
<th>Pitch repeated</th>
<th>Pitch alternated</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>447</td>
<td>472</td>
<td>455</td>
<td>441</td>
</tr>
<tr>
<td>1000</td>
<td>455</td>
<td>464</td>
<td>469</td>
<td>471</td>
</tr>
<tr>
<td>2000</td>
<td>411</td>
<td>434</td>
<td>420</td>
<td>427</td>
</tr>
<tr>
<td>3000</td>
<td>511</td>
<td>525</td>
<td>515</td>
<td>516</td>
</tr>
<tr>
<td>4000</td>
<td>480</td>
<td>496</td>
<td>482</td>
<td>483</td>
</tr>
</tbody>
</table>
To analyze the temporal dynamics of bindings, we again calculated partial repetition costs ($RT_{\text{partial repetition}} - RT_{\text{complete repetition/alternation}}$) for each effect type and RSI, and ran an ANOVA with RSI and effect type (color-pitch, color-response, and pitch-response). The only reliable effect was a main effect of the effect type, $F(2,100)=22.12$, $p<.001$, indicating higher costs associated with color-response integration (41 ms) than with pitch-response integration (9 ms) and color-pitch integration (3 ms), as can be seen in Figure 4.3.

![Figure 4.3](image)

**Figure 4.3.** Partial repetition costs as a function of response-stimulus interval (RSI) in Experiment 2.

We also compared findings across the two experiments by running a mixed-factor ANOVA on the partial repetition costs from Experiment 1 and 2 with RSI, effect type (relevant by irrelevant stimulus feature, relevant stimulus feature by response, and irrelevant stimulus feature by response), and experiment as factors. Experiment produced a main effect, $F(1,100)=14.91$, $p<.001$, due to higher average costs in Experiment 1 (34 ms) than in Experiment 2 (17 ms), and was involved in a two-way interaction with type, $F(2,200)=6.40$, $p<.005$. The
interaction reflected the pattern that is obvious from comparing Figures 4.2 and 4.3: The costs associated with the relevant stimulus feature and the response are comparable across the two experiments (38 vs. 40 ms), while the two types of costs involving the irrelevant stimulus feature (pitch in both cases) differ rather drastically. In fact, costs due to stimulus-stimulus interactions are more than 10 times larger in Experiment 1 than in Experiment 2 (35 vs. 3 ms), and the costs associated with pitch-response interactions are still about three times larger (28 vs. 9 ms).

This pattern is consistent with the idea that binding effects are mediated by the attentional set, and extend it to the apparently modality-specific allocation of attention. Particularly diagnostic are the pitch-related effects: The same task-irrelevant auditory feature that in auditory task interacts with other features and responses no less than relevant auditory features do, plays an only very minor role in a visual task.

**General Discussion**

In the present study, we aimed to obtain insights into the feature integration mechanisms by examining temporal dynamics of unimodal and multimodal bindings as indicated by partial repetition costs. Experiments 1 and 2 confirmed and extended previous observations of the spontaneous integration in and across multiple modalities: our findings suggest integration between auditory stimulus features (loudness and pitch), multimodal stimulus features (pitch and color), and between stimulus (unimodal or multimodal) and response.

With respect to the first question that guided our study, whether the decay rates for bindings involving visual and auditory feature were comparable, our findings suggest a clear-cut conclusion. For one, the RSI functions (see Figure 4.2) obtained in the unimodal auditory Experiment 1 look very similar to the time functions reported by Hommel and Colzato (2004), with some decrease over time and first indications of decreasing stability after about four seconds. Moreover,
the RSI functions of loudness-response interactions in Experiment 1 and of color-response interactions in Experiment 2, the two interactions involving the relevant stimulus feature and the also task-relevant response, look very similar as well, with particularly high partial repetition costs at short RSIs and an asymptote around 3000 ms. At first sight, the effects indicative of the integration of stimulus features seem to show marked differences, as the unimodal auditory bindings in Experiment 1 are much more pronounced than the audiovisual bindings in Experiment 2. However, note that the corresponding RSI functions are extremely similar (as also confirmed by the absence of a RSI-by-experiment interaction in the analysis of partial repetition costs), suggesting that the temporal dynamics of the two effects are comparable. In other words, the differences lay in the intercept (a higher departure level at short RSIs in the unimodal case) but not in the slope (the decrease of effect sizes with increasing RSI) of the binding cost-by-RSI function.

Our second guiding question concerned the role of attention both in unimodal and multimodal feature binding. We examined whether task relevance of the features involved would affect the temporal dynamics of the bindings. Experiment 1 did not provide evidence for this possibility. Even though a close-to-significant effect of type indicated that pitch-response interactions were somewhat less pronounced than loudness-response interactions, the difference was not large and did not interact with RSI. In fact, the RSI functions of all three binding-related effects are not only almost the same in terms of intercept but virtually identical in terms of slope (see Figure 4.2). So, whatever causes the decrease of effect size over time it affects task-relevant and task-irrelevant features in the same way.

Our third question concerned the impact of the context or attentional set on uni- and multimodal bindings and their temporal dynamics. Particularly diagnostic was the interaction between pitch repetition and response repetition, which was observed in both experiments. A comparison between Figure 4.2 and
Figures 4.3 shows that the answer is somewhat complicated by what seems to be a mere floor effect (in Experiment 2) but rather clear. The intercept of the interaction differs in the two experiments, with a much higher starting level in the unimodal Experiment 1 than in the multimodal Experiment 2. Given the much lower starting point of the size of the interaction in Experiment 2, it is not surprising that the effect hits base rather quickly and stays around zero from the second RSI on. That is, the lower overall impact of bindings including irrelevant features in Experiment 2 as compared to Experiment 1 does not seem to be due to faster decay but, rather, to a stronger point of departure.

Taken altogether, the results of our study point to two, apparently, general principles that seem to regulate the temporal dynamics of binding-related effects. The first one is task relevance. Earlier observations have repeatedly suggested that binding-related effects are more likely and more reliable if the features they involve are relevant to the task at hand (and thus attended), such as if they signal the response to S2 (Hommel, 1998; Hommel & Colzato, 2004; Zmigrod & Hommel, 2009). Indeed, increasing task demands by turning a simple detection task into a discrimination task has been shown to induce the allocation of more attentional resources (Luck & Hillyard, 1994, 1995; Treisman & Sato, 1990) and to elicit a larger N2pc (Hyun, Woodman, Luck, 2009)—an ERP component that is taken to reflect the allocation of visual attention. That the allocation of more resources boosts binding effects is consistent with observations of Talsma and Woldorff (2005), who found larger audio-visual integration effects in the ERP for attended than for unattended stimuli.

However, task relevance may also be more implicit and generalize from response to stimulus features. For instance, Hommel (2007b) investigated why location features are typically involved in stimulus-related features and stimulus-response-related bindings even though the location of S1 or S2 is commonly not relevant in sequential-effect tasks. One possibility is that defining responses in terms of left and right, which they commonly are in such tasks, makes the
location task relevant. If one further considers that action control operates on representations of perceptual action effects (Hommel, 1996, 2009), so that selecting between actions considers the perceived location of these actions, controlling spatially defined actions may require attending to location no less than selecting a color-defined target requires attending to color. These attentional requirements may lead to a preparatory priming of all the feature dimensions involved, so that features coded on these dimensions are weighted more strongly (Hommel, 2004, 2010). Theoretically, this weighting may affect both the creation of bindings and their retrieval induced by repeating features. However, given the far-reaching automaticity of the integration process (Hommel, 2005, 2007b; Hommel & Colzato, 2004) and hints to a selective impact of attentional manipulations on retrieval (Hommel, Memelink, Zmigrod & Colzato, submitted), feature weighting may mainly or exclusively control the retrieval of bindings.

Our present findings suggest that feature weighting does not hinge on the task relevance of the respective feature dimensions alone but also on the sensory modalities involved. Take, for instance, the impact of pitch-response bindings in the two experiments. The mere fact that the relevant feature was auditory in Experiment 1 was apparently sufficient to boost this impact by a factor of three as compared to Experiment 2, where the relevant stimulus was visual. This suggests that defining a sensory modality as task relevant induces an attentional set that potentiates the impact of stimuli processed by this modality, irrespective of whether the dimension of these stimuli is relevant or irrelevant for the task. To put it differently, stimulus features seem to be weighted according to any kind of match with whatever the cognitive system considers relevant (Pratt & Hommel, 2003). If the system is prepared to process auditory loudness, feature codes are potentiated to the degree that they refer to loudness, to the auditory modality, or both, suggesting that modality is treated just like any other feature that does or does not match with a currently processed stimulus. According to this principle, all three “features” combined in Experiment 1 were task relevant: loudness
because it signaled R2, the response (location) because it needed to be selected, and pitch because it belonged to the task-relevant stimulus modality. In Experiment 2, color and response (location) were also relevant but pitch no longer was. With regard to the intercept-related effect pattern, this principle accounts for all our main observations: all interactions involving a task-relevant feature as just defined were sizable and rather robust across most of the RSI, whereas the interactions including a feature that did not fall under this definition of task relevance were weak and restricted to the shortest RSI.

The other general principle that is visible in our data pattern refers to what we up to now have called “decay”. Whatever the point of departure, binding-related effects disappear over time. How quickly they disappear obviously depends on the departure level, that is, on how strong the corresponding effect is at short RSI, but we had no indication that the rate of disappearance would be systematically related to this level or to the task relevance or modality of the features involved. In other words, the impact of features on binding, or the retrieval of bindings, is regulated through a mechanism that operates on the intercept but not the slope of the effect-time functions. If we attribute intercept effects to the weighting of feature dimensions or modalities in the light on their task relevance, how is the apparently invariant slope effect to be explained?

The probably most obvious account was already suggested by our use of the term “decay”. Neural codes generally increase and then decrease in activation and so cognitive codes have often been assumed to be subject to decay, that is, to a spontaneous decrease of activation over time. To account for negative slopes in our time functions, two types of decays may be considered. One refers to the present activation state of a given binding. Bindings may be created by linking the feature codes that were activated by the stimulus, and this link may decay after some time. This picture would be consistent with the original idea underlying the concept of an “object file” (Kahneman et al., 1992), which has been assumed to mediate the perception of object constancy by bridging brief intervals in which an
object disappears from view. Object files would link the feature codes referring to an object, but it would only do so for a certain amount of time, namely, as long as object constancy is given. In other words, object files should have a limited lifetime that should correspond to the interval that still allows for perceiving an object as constant. In the original study (Kahneman et al., 1992) the delay between two presentations of the critical stimuli was in the range of milliseconds, yet later studies found that object files can persist as long as eight seconds (Noles, Scholl, & Mitroff, 2005). It may be interesting to note that these rather long lifetimes were observed under conditions where the interval between two presentations was filled by other visual events, events that served to bridge between the two presentations by implying some kind of change (such as the rotation of the display, which contained placeholders of the actual stimuli). In contrast, in our study the stimuli simply disappeared and the interval between S1 and S2 was void of any other perceptual event (except the participant's own response). From studies on causal perception is known that intermediate, bridging events extend the operation space of events, that is, the length of the interval across which two events are perceived to be related to each other (e.g., Reed, 1992, 1999; for overviews, see Einhorn & Hogarth, 1986; Gruber, Fink, & Damm, 1957). With this in mind, our observation of binding lifetimes of about four seconds or so does not seem to be inconsistent with findings of longer lifetimes in studies using bridging events.

Alternatively, decay may not represent the activation state of a feature link or object file but to the actual or functional lifetime of an episodic memory trace. This trace may or may not be active after having been created, but it may be either deleted after some time or alternatively may be re- or over-written after a new combination of features appeared (Alvarez & Thompson, 2009) (which would refer to its actual lifetime) or at least no longer be discriminable from alternative traces (which would refer to its functional lifetime). Note that our analyses focus on two particular representations (of S1 and S2) in which the features in question
are combined in a particular way, but more than hundred combinations are coded in a typical sequential-effects experiment. If each single combination would be stored over longer time (and there is some evidence that this is an actual possibility: DeSchepper & Treisman, 1996; Logan, 1988), each single trace would compete with an increasing number of other traces, which again would make it increasingly difficult to discriminate it from these alternatives. The difficulty of discrimination may increase over time, so that the most recent object file may no longer “stick out” after about four seconds. Given that all combinations of features are equally frequent and probable in a typical sequential-effects task, the selective impact of the most recent combination would thus “wash out” over time.

A third alternative also refers to temporal discriminability, but from a slightly different angle. If object files, and in fact any feature binding, indeed serve to relate different events to each other (e.g., two successive views of the same object), it makes sense to assume that there are certain limitations to this bridging functions. Under realistic circumstances, a certain time where an object remains invisible may be tolerable with respect to perceived object constancy or persistency, but at some point this tolerance should end. Indeed, there is evidence that spatiotemporal continuity is crucial for object persistency (Gao & Scholl, 2010; Mitroff, & Alvarez 2007; Yi et al., 2008) and that the impression of a continuously existing object begins to fade if visual objects disappear for more than half a second (Burke, 1952)—at least in the absence of bridging events (see above). This implies that there might be a criterion for relating two temporally separated events, a criterion that determines whether the events are integrated or segregated. These criteria seem to depend on the temporal density of events (Akyürek, Toffanin, & Hommel, 2008), which in view of the wide temporal spread of the events in our setup implies a rather lax criterion that may well fit with the disappearance of effects around 3-4 seconds. It may be such a criterion that decides whether the previous binding (coding S1 and R1) is related to the present stimulus-response episode (S2/R2), or that at least modulates the retrieval
of the previous binding. If so, extended blank intervals between the S1/R1 and S2/R2 events may mainly serve to signal that these events are independent, which might decrease the likelihood that the previous episode is retrieved.

A fourth alternative might be that memory plays a role in the longer RSIs. Hommel and Colzato (2004) have provided some evidence that the integration of visual features takes about several hundred milliseconds, and one may imagine that the RSI manipulation taps into an extended process comprising of the successive coding, integration, and consolidation of feature information, which eventually results in the creation of a long-term memory trace. The consolidation of a binding and/or the creation of such a trace may somehow make the bound information less accessible and/or prevent the retrieval of the previous binding, which would account for the decay rates we observed. Even though this is a possible account, we note that previous investigations on the relationship between memory processes and binding did not show hints to an interaction. For instance, Colzato, Raffone, and Hommel (2006) failed to find stronger binding effects for over learned feature combinations; for instance, the binding between the picture of a banana and of the color yellow was no stronger between the banana and the color purple (see also Hommel & Colzato, 2009). Saiki (2009) was equally unsuccessful in finding an impact of memory on binding with a visual search task.

At this point, the available data do not allow for discriminating between these possibilities and the accounts are not necessarily mutually exclusive either. That is, the probability of retrieving a feature binding may well be co-determined by a number of factors, including the task relevance of the features involved, spontaneous decay of their binding, the temporal discriminability between the binding and previous, alternative bindings, and the width of the temporal integration window according to which the previous event (S1/R1) and the present event (S2/R2) are related. As far as our observations suggest, the scenario holds for stimuli processed by different modalities and for both unimodal and multimodal events.
Footnotes

1. Here we use the terms “decay” and “decay rate” in a rather descriptive fashion, merely to capture the observation of a decrease of effect sizes over time (RSI). The possible reasons for that (actual decay of the bindings, increasing temporal discrimination, etc.) will be discussed in more detail in the General Discussion.

2. Partial repetition costs for a given interaction between factors X and Y were calculated as the difference between the RTs for partial repetitions (feature X repeated and feature Y alternated, or vice versa) and the RTs for complete repetitions and “complete” alternations. E.g., the partial repetition costs for the loudness X pitch interaction at a given RSI would be PRCloudnessXpitch = (RTloudness repeated/pitch alternated + RTloudness alternated/pitch repeated)/2 - (RTloudness repeated/pitch repeated + RTloudness alternated/pitch alternated)/2. Partial repetition costs thus correspond to the 2-way interaction term of the respective features (and are thus immune to possible, but theoretically less relevant, main effects of feature repetition); a value close to zero mean that the repetition effects of the two given features do not interact; a value greater than zero indicates a “binding-type” interaction of the sort described in the text.
Chapter 5

Cognitive Flexibility and Control in Children with Autistic Spectrum Disorder

Abstract

Autistic Spectrum Disorder (ASD) has been claimed to be associated with impaired cognitive flexibility, but the evidence is equivocal. We compared 33 ASD-diagnosed and 33 normally developing children in the age of 10-18 in a task that assesses the integration and updating of stimulus-response episodes. Children with ASD showed more, rather than less pronounced aftereffects of integration, suggesting that they are not impaired in binding stimulus and response features but in updating bindings that are no longer valid. This impairment was correlated with the lack of flexibility in a task-switching context but not with an index of inhibitory control. The findings are taken to provide evidence for a specific impairment of cognitive flexibility in ASD, presumably due to prefrontal dopaminergic hypoactivity.
Introduction

Autistic Spectrum Disorder (ASD) is one of the most common childhood disorders, and characterized by social communication impairment, deficits in language skills, and repetitive behaviors. Various authors have advocated various factors that might account for the disorder, but it is fair to say that there is a rather general agreement that impairments related to executive control functions play a major role (for reviews, see Corbett, Constantine, Hendren, Rocke, & Ozonoff, 2009; Hill, 2004; Kenworthy, Black, Harrison, Della Rosa, & Wallace, 2009). Among other things, these impairments are assumed to render ASD patients cognitively less flexible, which would account for both impaired performance in clinical tests like the Wisconsin card sorting task (Willcutt, Sonuga-Barke, Nigg, & Sergeant, 2008) and behavioral rigidity in everyday life behavior.

Unfortunately, however, experimental evidence supporting the link between ASD and cognitive flexibility is still scarce and equivocal (Geurts, Corbett, & Solomon, 2009). As suggested by Geurts et al., this might be due to the fact that most clinical tests are rather complex and unlikely to provide process-pure measures of the interesting cognitive processes. For instance, the Wisconsin card sorting task relies on a good understanding of the task, working memory, learning from feedback, the availability of multiple strategies, and so on, and not all of these abilities and skills are related to the processes targeted by executive-control and flexibility accounts of ASD, thus the outcome on these kinds of test is multi-interpretable. Therefore, there is a need for more diagnostic experimental tasks that provide more process-pure measures of cognitive flexibility.

In the present study, we considered one such task that is a rather well understood with regard to its neural (Kühn, Keizer, Colzato, Rombouts & Hommel, in press) and neuromodulatory (Colzato & Hommel, 2008) basis and its theoretical implications (Hommel, 2004), and that has been shown to be sensitive
to individual differences in fluid intelligence (Colzato, van Wouwe, Lavender, & Hommel, 2006) and age (Hommel, Kray, & Lindenberger, submitted). As we will describe, this task assesses the individual ability to handle episodic bindings of feature codes representing objects and sensorimotor events (so-called event files: Hommel, 1998), a process that is likely to capture at least one aspect of the cognitive impairment expressed in ASD. We thus pursue an analytical, piecemeal approach that does not try to assess and explain the whole disorder at once, but rather attempts to identify selected, important aspects of the disorder by using a relatively simple, well-defined experimental task that taps into a set of relatively well-understood low-level processes.

Let us first introduce the task and its theoretical background. Given that the primate cortex processes the various features of perceptual events and actions in distinct brain regions (e.g., DeYoe & van Essen, 1988), it has been assumed that the representations of these features need to be integrated into coherent episodic bindings (e.g., Hommel, 2004; Kahneman, Treisman, & Gibbs, 1992). Evidence for the spontaneous integration of perceptual features comes from analyses of (interactions between) repetition effects. For instance, people not only respond faster to letters that they just saw in a preview display (a standard priming or repetition effect), but they are particularly fast if the repeated letter also appears in the same location (Kahneman et al., 1992). This suggests that processing a perceptual event induces the binding of the codes of its features (e.g., letter shape and location in the given example), so that repeating the particular conjunction of features allows for particularly efficient processing. Comparable observations have been made for auditory features (Mondor, Hurlburt, & Thorne, 2003; Zmigrod & Hommel, 2009), perceptual features from different sensory modalities (Zmigrod, Spapé, & Hommel, 2009), and for perceptual and action features (Hommel, 1998). For instance, responding to object A by carrying out response X is easier after having paired A and X, or the unrelated object B and action Y, than after having responded differently to the same object (A→Y) or
responded similarly to a different object (B→X). Apparently, then, a single pairing of a stimulus (feature) and a response is sufficient to create an episodic binding (an event file; Hommel, 1998) that interferes with partially, but not completely overlapping bindings. This suggests that repeating at least one (stimulus and/or response) feature leads to the retrieval of the just created binding, which interferes with current processing if that involves the reactivation of a no longer valid feature code (Hommel, 2004).

Relating these observations to the possible impairments underlying ASD, one might hypothesize that is the integration of features (the binding process) that is impaired in these populations (e.g., Frith, 2003). If so, one would expect that partial-repetition costs (i.e., the performance deficits with incomplete repetitions of stimulus-feature or stimulus-response combinations as compared to complete repetitions or alternations) are less pronounced with autism-related disorders. However, there is another perhaps more plausible possibility. Note that partial-repetition costs can only be observed if two conditions are met: feature codes need to be integrated in the respective prime trial; and this created binding needs to be retrieved in the present (probe) trial. Interestingly, attempts to dissociate these two processes provided evidence that the binding process proper is more or less automatic (Hommel, 2005), whereas the retrieval process is affected by task instructions and individual differences, suggesting at least some degree of control. For instance, bindings involving task-relevant features have a stronger impact on behavior (Hommel, 1998, 2007b), suggesting that they are more likely to be retrieved. Additionally, partial-repetition costs are more pronounced in individuals with low fluid intelligence (Colzato et al., 2006), and in young children and elderly participants, as compared to young adults (Hommel, Kray, & Lindenberger, submitted). Given that executive-control functions are related to fluid intelligence (Duncan et al., 2000), not fully developed in young age (Hongwanishkul, Happaney, Lee, & Zelazo, 2005), and impaired in old age (Fisk & Sharp, 2004), these observations suggest that more efficient control functions
are reducing the impact of previously created feature bindings, presumably by (better) restricting memory retrieval to the task-relevant information. This is also consistent with recent findings of Keizer, Verment, and Hommel (2010), whose participants received neurofeedback to increase cortical gamma synchronization. Such training improved memory retrieval in a standard recollection task and reduced partial-repetition costs.

In view of these hints to a link between executive control functions and the management of episodic event files, we hypothesized that ASD is associated with impairments in this management and, thus, in the flexibility of assessing and switching between episodic representations. This would fit with the observation of specific deficits in children suffering from ASD in tasks requiring mental flexibility, such as set-shifting tasks or the Wisconsin card sorting task (for reviews, see Corbett et al., 2009; Gioia, Isquith, Kenworthy, & Barton, 2002; Kenworthy et al., 2009). If ASD would indeed be associated with poorer event-file management abilities, one would expect ASD patients to show more pronounced partial-repetition costs than control participants do.

We tested this hypothesis by comparing the performance of a group of ASD-diagnosed children and a group of normally developing children in a standard event-file task (e.g., Hommel, 1998), which was only slightly adapted for the use with children. As a converging measure, we also included the set-shifting task from the Amsterdam Neuropsychological Tasks (ANT: de Sonneville, 1999), which provides indices for two types of executive functions: flexibility and inhibition of response sets. This task has successfully been used in participants with impaired frontal functioning and attention problems (e.g. de Sonneville et al., 2002; Huijbregts, de Sonneville, Licht, Sergeant, & van Spronsen, 2002).
Method

Participants
The autism spectrum disorders' participants (ASD) were selected from consecutive referrals to the outpatient and inpatient department of child and adolescent psychiatry at the University Medical Centre of Utrecht, the Netherlands. Two certified experienced child psychiatrists diagnosed these participants using DSM-IV criteria (American Psychiatric Association, 1994). The group included 33 participants (23 males) between 11 and 18 years of age (mean=14, SD=2.2). The control group (typically developing participants) comprised of 33 healthy participants (26 males) between 10 and 18 years of age (mean=15, SD=2.2). There was no significant difference in age between the groups. All participants had full scale IQ above 70, as measured with the Dutch adaptations of the Wechsler Intelligence Scale for Children (Wechsler, 1997). All participants reported having a normal or corrected-to-normal vision.

Procedure
The study was conducted in accordance with the declaration of Helsinki and guidelines of the local ethics committee. All parents signed a written consent before participating in the study.

Event file task
The event file task measures binding-related effects by diagnosing partial-repetition costs related to combinations of stimulus features (shape and color in our case) and combinations of stimulus features and the response. To manipulate the repetition versus alternation of stimulus features and responses, the task comprises of pairs of trials with a prime trial (S1→R1) followed by a probe trial (S2→R2), see Figure 5.1. The probe trial required a manual binary-choice response (R2) to the shape of the second stimulus S2 (an apple or a banana). The prime trial required a manual response (R1) to the mere onset of the first stimulus
The correct R1 was signaled in advance of S1 (through a left- or right-pointing arrowhead), so that S1 and R1 could be varied independently, which was necessary to create orthogonal repetitions and alternations of stimulus shape and response. As an additional stimulus feature, color was also varied by presenting the apple or banana in green or yellow (see Colzato, Raffone, & Hommel, 2006). Stimulus color could repeat or alternate independently of stimulus shape and responses, thus creating a 2x2x2-factorial design.

The experiment was composed of a practice block with 10 practice trials, which were not further analyzed, and an experimental block with 196 experimental trials. The order of the trials was randomized but all eight conditions appeared equally often. Half of the participants responded to the apple and the banana by pressing on the left and right key press, respectively, while the other half received the opposite mapping. The participants were asked to respond as quickly and accurately as possible.

**Figure 5.1.** Sequence of events in the event file task. A visual response cue signaled a left or right response (R1) that was to be delayed until presentation of the first stimulus S1 (S1 is used as a detection signal for R1). The second stimulus S2 appeared 1000 ms after S1. S2 signaled R2, a speeded left or right response according to the shape.
Set-shifting task

The set-shifting task measures attentional flexibility and the ability to inhibit inappropriate habitual response tendencies by assessing performance in a task where participants need to switch between two competing response sets. A colored square moved randomly to the left and right on a horizontal bar consisting of 10 squares (see Figure 5.2). The task comprised of three parts and instructions were given before each part. Part 1 (fixed-compatible condition, 40 trials) required spatially compatible responses to the motion of the green-colored square: a click of the left mouse key if the square moved left and a right click if it moved right. Part 2 (fixed-incompatible condition, 40 trials) required directionally incompatible responses: a left click if the square moved right and a right click if the square moved left. In part 3 (random condition, 80 trials), the color of the square varied randomly between green and red. When the color of the square after the move was green, a compatible response was required, as in part 1. When the color of the square was red, an incompatible response was required, as in part 2. Thus, due to the unpredictability of the direction of the motion and the type of the task (compatible vs. incompatible), high levels of mental flexibility were required by continuously having to adjust the response rule.

Participants were to respond between 150 and 8000 ms after a signal onset, otherwise the trial was automatically replaced by a new trial. Next signal onset was always 250 ms after the response. Inhibitory control was measured by contrasting the performance in the fixed-incompatible condition (part 2) with the fixed compatible condition (part 1). Flexibility was measured by contrasting the performance in the random compatible condition (part 3) with the performance in the fixed compatible condition (part 1).
Figure 5.2. Set-shifting task: Timing between signals and an example of two consecutive trials in part 3 of the task. In trial (i+1) the block has jumped to the left and has turned green: the correct response is to press the left button (compatible response). In trial (i+2) the block has jumped to the left and the color changed to red: the correct response is now to press the right button (incompatible response). PRI=Post-Response Interval.

Results

Event file task

Trials with incorrect R1 responses (0.8%), as well as missing (RT>1500 ms) or anticipatory (RT<100 ms) R2 responses (0.02%) were excluded from analysis. The mean reaction time for correct R1 was 446 ms (SD=220). From the remaining data, mean reaction time (RTs) and percentage of errors (PEs) for R2 (see Table 5.1) were analyzed as a function of the four variables: the relationship between S1 and S2 (repetition vs. alternation) with regard to shape and to color, and the relationship between responses R1 and R2 (repetition vs. alternation), which all varied within participants, and group (ASD vs. control). Mixed-design ANOVAs were performed with repeated measures on three variables and with group as between-participant variable.
Table 5.1. Event file task: Means of mean reaction times for responses to stimulus 2 ($RT_{R2}$ in ms) as a function of group (ASD children vs. control - typically developing children), the relationship between the responses (R1 and R2), and the relationship between the stimuli features (S1 and S2) for shape and color. The rightmost column gives the partial repetition costs (see FOOTNOTE 1), which differed significantly in response-shape between the two groups, $p<.005$, both in reaction times and error rates.

<table>
<thead>
<tr>
<th>Group</th>
<th>Response repeated</th>
<th>Response alternated</th>
<th>Partial repetition costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Shape repeated</td>
<td>Shape alternated</td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>518</td>
<td>649</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>576</td>
<td>635</td>
<td></td>
</tr>
<tr>
<td>Errors (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>5.7</td>
<td>18.1</td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>1.3</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td>RTs (ms)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>580</td>
<td>575</td>
<td>8</td>
</tr>
<tr>
<td>Control</td>
<td>600</td>
<td>592</td>
<td>5</td>
</tr>
<tr>
<td>Errors (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td>11.4</td>
<td>10.2</td>
<td>1.2</td>
</tr>
<tr>
<td>Control</td>
<td>3.6</td>
<td>7.3</td>
<td>1.8</td>
</tr>
</tbody>
</table>

The groups (ASD vs. control) did not differ in RTs but the ASD group showed more errors (10.7%) than the control group (4.7%), $F(1,64)=22.2$, $p<.0001$. There was also a significant main effect of shape repetition in RTs, $F(1,64)=25.37$, $p<.0001$, due to faster responses to repeated (575 ms) than alternated shapes (600 ms). This effect was modified by group, $F(1,64)=15.36$, $p<.0001$, due to a more pronounced shape-repetition effect in the ASD group. In the error rates, response repetition interacted with group, $F(1,64)=4.51$, $p<.05$, whereas the control group exhibited a response-repetition benefit (with 3.7% and 5.6% errors in repetition and alternation trials, respectively), the ASD group showed the opposite pattern (11.9% vs. 9.5%).
There were significant interactions between shape repetition and response repetition in RTs, $F(1,64)=164.92$, $p<.001$; and PEs, $F(1,64)=109.37$, $p<.0001$. These findings followed the common pattern with worse performance if only one of the features (shape or response) is repeated while the other is not, as compared to complete repetitions or alternations (see Hommel, 1998). In addition, there was a significant interaction between color, the irrelevant feature, and the response in PEs, $F(1,64)=9.14$, $p<.005$. This interaction was further modified by shape repetition, $F(1,64)=7.49$, $p<.01$, due to particularly accurate performance if all three features were either repeated or alternated—a common pattern that has been attributed to shortcutting response selection processes with complete repetitions (Bertelson, 1963) and alternations (Hommel & Colzato, 2004).

More importantly for our study, the response-shape interaction was further modified by group in both RTs, $F(1,64)=8.98$, $p<.005$; and PEs, $F(1,64)=8.37$, $p<.005$. This was due to more pronounced interactions in children with ASD than in typically developing children. In contrast, group was not involved in either the three-way interactions with shape and color or with color and response, all $Fs(1,64)<1$, or the four-way interaction, $F(1,64)=1.09$, $p=.30$, and $F(1,64)=1.95$, $p=.17$, for RTs and PEs, respectively.

**Set-Shifting task**

In the analyses of the set-shifting task, the data for one control and three ASD participants were lost due to technical problems. As an index of flexibility, we calculated the difference in performance, both in RT and PE, in the random compatible condition (of part 3) and the fixed compatible condition (of part 1)—so that higher scores indicate less flexibility. As an index of inhibition, we calculated the difference in performance in the fixed incompatible condition (part 2) and the fixed compatible condition (part 1)—so that higher scores indicate less efficient inhibition.
T-tests revealed that the flexibility index discriminated between the two groups in RTs, $t=2.17$, $p=.034$, but not in PEs, $t=1.64$, $p=.11$, whereas no significant effects were obtained for the inhibition index in either RTs, $p=.10$, or PEs, $p=.13$.

**Correlations**

We quantified the three binary partial-repetition effects (shape/color, shape/response, and color/response) by calculating the interaction terms ($RT/PE_{\text{partial repetition}}-RT/PE_{\text{complete repetition/alternation}})/2$ (FOOTNOTE 1), see Table 5.1, and correlated these measures with the corresponding flexibility and inhibition indices from RTs and PEs.

The flexibility index correlated with shape-response partial-repetition costs in RTs, $r=.29$, $p<.05$, and marginally so in PEs, $r=.23$, $p=.07$, indicating that less flexibility was associated with more pronounced partial-repetition costs. The flexibility index did not correlate with either shape-color or color-response partial-repetition costs (all $ps>.26$). The inhibition index was not involved in any reliable correlation with partial-repetition costs, all $ps>.14$. The two indices correlated only mildly in RTs, $r=.22$, $p=.08$, and not at all in error rates, $r=.14$, $p>.28$.

**Discussion**

The aim of this study was to investigate whether children suffering from ASD would show a specific effect in a task tapping into the handling of episodic event files, that is, of bindings between codes that represent the features of experienced objects and stimulus-response episodes (Hommel, 1998, 2004). Both normally developing controls and the ASD group showed partial-repetition costs for combinations of the two task-relevant stimulus and response features: stimulus shape and response location. It is known that task relevance modulates feature-integration effects (e.g., Hommel, 1998; Hommel & Colzato, 2004), so that it is
not surprising that reliable effects were mainly restricted to the features that mattered for the task. More interesting, however, is that these effects were observed in both groups. Given that partial-repetition costs in S2-R2 performance can only occur if the respective features and be integrated when processing S1 and R1, this observation implies that binding as such does not seem to be impaired in ASD. This does not seem to fit with the claim that ASD is associated with difficulties in feature integration (e.g., Frith, 2003).

Importantly, the aftereffects of binding were rather larger than smaller in the ASD group, suggests that ASD impairs the handling of bindings. That is, both healthy controls and children suffering from ASD seem to spontaneously integrate stimulus-response episodes and automatically retrieve traces of these episodes when facing a similar, that is, feature-overlapping episode thereafter. However, healthy controls seem to be more efficient in preventing these traces from affecting ongoing processes if they do not fit with the current feature combinations. One may consider two ways in which that might be done. For one, healthy controls may be better in inhibiting retrieved but no longer valid traces. However, we have seen that partial-repetition costs were correlated with the flexibility index but not the inhibition index, suggesting that inhibition did not play a major role in producing or reducing these costs. Another possibility is that healthy controls are more efficient in updating feature bindings, that is, in replacing retrieved but no longer valid bindings by new bindings. This would fit with the observation that higher between-repetition costs were accompanied by lower flexibility (i.e., by higher scores in the flexibility index). Hence, there are reasons to assume that repetition costs provide a relatively pure measure of flexibility, at least with respect to the updating of cognitive representations.

Further evidence for an interesting theoretical and empirical link between partial-repetition costs, control functions, and ASD comes from research on the neuromodulation of the underlying cognitive processes and in particular from the fact that they all seem to rely on prefrontal dopaminergic pathways. ASD is
considered a “hyperdopaminergic” disorder along with ADHD and schizophrenia (Previc, 2007). Considering the comorbidity between ASD and ADHD, it seems plausible that a dopamine-related abnormality is the common source for the similar symptoms in these two disorders (Gillberg & Billstedt, 2000). In addition, administrating risperidone (a dopamine-receptor antagonist) reduces some of the behavioral symptoms in ASD children (McCracken et al., 2002). Moreover, ASD considered as one of the most highly heritable developmental disorder, and a number of genes linked to it (see: Yonan et al., 2003) are associated with dopamine, such as DBH (Robinson et al., 2001). In addition, the dopamine transporter (DAT1) genotype, which is associated with ADHD, tics and anxiety found in ASD population (Gadow, Roohi, DeVincent, & Hatchwell, 2008).

The dopaminergic system is also involved in a number of executive control functions, such as planning, working memory, or temporal sequencing (for a review, see Previc, 1999). More relevant to our study, there is compelling evidence that dopaminergic system is important to mental flexibility and cognitive shifting operations. For instance, older adults show declines in dopaminergic transmission related to D1 (Rinne, Lonnberg, & Marjamaik, 1990; Suhara et al. 1991) and D2 receptors (Rinne et al., 1990; Volkow et al., 1996), and these declines are associated with poor performance in many neuropsychological control-related tests, such as the Stroop task, the Wisconsin sorting card task, and others (Volkow et al., 1998). This decline in cognitive ability can be corrected by administrating dopaminergic agonist such as Pirebidil (Ollat, 1992). Likewise, Braver et al. (2001) demonstrated that aged adults are impaired in proactive control, which is associated with the dopamine level in the PFC. Furthermore, flexibility is improved by inducing positive affect (Dreisbach, 2006; Van Wouwe, Band, & Ridderinkhof, in press), which is assumed to induce temporary increases of the dopamine level (Ashby, Isen & Turken 1999;
Dreisbach & Goschke, 2004). Thus, cognitive control and flexibility in particular appear to be modulated by dopamine.

Given that both ASD and cognitive flexibility seem to depend on dopaminergic pathways, it is interesting that the same seems to be true for partial-repetition costs. For instance, these costs are systematically modulated by affect (Colzato, van Wouwe, & Hommel, 2007a) and related to individual differences in the spontaneous eyeblink rate (Colzato, van Wouwe, & Hommel, 2007b), a clinical marker for the level of dopaminergic functioning (Blin, Masson, Azulay, Fondarai, & Serratrice, 1990; Kleven & Koek, 1996). Moreover, aftereffects of stimulus-response bindings are affected by stress (Colzato, Kool, & Hommel, 2008) and the use of cannabis but not cocaine (Colzato & Hommel, 2008), suggesting that it is mainly dopaminergic D1 receptors that are involved but not D2 receptors. Given that D1 but not D2 receptors are dominant in the mesocortical dopaminergic pathways, which are also assumed to drive executive control functions including working memory (e.g., Arnsten, & Goldman-Rakic, 1998), these observations provide converging evidence for a link between ASD, executive control, and the management of episodic feature bindings.

To conclude, the present study provides evidence that ASD is associated with a specific deficit in updating episodic stimulus-response representations. The degree of this deficit is correlated with the lack of flexibility in task-switching performance, which suggests that even the relatively simple task we used to assess aftereffects of feature binding captures the essence of processes that also impair performance in more complex experimental tasks and neuropsychological tests.
Partial-repetition costs for a given interaction between factors X and Y were calculated as the difference between the RTs/PEs for partial repetitions (feature X repeated and feature Y alternated, or vice versa) and the RTs/PEs for complete repetitions and “complete” alternations. E.g., the partial repetition costs in RTs for the shape X response interaction at a given group would be PRCshapeXresponse = (RT shape repeated/response alternated + RT shape alternated/response repeated)/2 - (RT shape repeated/response repeated + RT shape alternated/response alternated)/2. Partial repetition costs thus correspond to the 2-way interaction term of the respective features (and are thus immune to possible, but theoretically less relevant, main effects of feature repetition); a value close to zero mean that the repetition effects of the two given features do not interact; a value greater than zero indicates a “binding-type” interaction of the sort described in the text.
Chapter 6

The Relationship between Feature Binding and Consciousness: Evidence from Asynchronous Multimodal Stimuli

Abstract

Processing the various features from different feature maps and modalities in coherent ways requires a dedicated integration mechanism (“the binding problem”). Many authors have related feature binding to conscious awareness but little is known about how tight this relationship really is. We presented subjects with asynchronous audiovisual stimuli and tested whether the two features were integrated. The results show that binding took place up to 350 ms feature-onset asynchronies, suggesting that integration covers a relatively wide temporal window. We also asked subjects to explicitly judge whether the two features would belong to the same or to different events. Unsurprisingly, synchrony judgments decreased with increasing asynchrony. Most importantly, feature binding was entirely unaffected by conscious experience: features were bound whether they were experienced as occurring together or as belonging to a separate events, suggesting that the conscious experience of unity is not a prerequisite for, or a direct consequence of binding.
**Introduction**

We perceive the world through several sensory modalities and process the numerous features of the events we perceive in various cortical maps (e.g., Zeki & Bartels, 1999; Kaas & Hackett, 1999). Many authors have noted that these processing characteristics are likely to create all sorts of binding problems: how does the brain know which of the currently coded features belong to the same event (von der Malsburg, 1999) and how do we integrate all these different features into one coherent conscious representation (Treisman, 2006). It is true that one can argue whether having a coherent conscious experience of a multi-featured event really requires the actual binding of the corresponding feature codes. After all, all these codes are located in the same brain and, if we assume that conscious states are lawfully related to brain states, this may be sufficient to guarantee coherence. And yet, given that we can process (though not necessarily attend to) several objects, and control multiple actions concurrently, a whole number of binding problems needs to be solved in any case. And given that our conscious experience is commonly restricted to only some, often just one of these objects and actions, consciousness is likely to rely on at least some form of feature binding.

Even though there is no really comprehensive theory of the relationship between feature binding and consciousness, several authors have claimed that these two processes are tightly related (for an overview, see Engel & Singer, 2001). For instance, Treisman (2003) assumes that feature integration is a necessary precondition for coherent conscious perception, and that focused attention is required and responsible for creating feature bindings. Along the same lines, Crick and Koch (1990) and Engel and Singer (2001) have claimed that feature binding, and the neural processes underlying it, is an essential precondition for conscious awareness. At the same time, however, there is increasing evidence that attention is not necessary for binding (see Hyun,
Woodman, & Luck, 2009) and that feature binding and conscious awareness, or the processes underlying them, can be dissociated. For instance, Wojciulik and Kanwisher (1998) observed that explicit feature binding (i.e., reporting the relationship between multiple features) is impaired in Balint’s syndrome while implicit feature binding is not. In healthy subjects, Mitroff, Scholl, and Wynn (2005) found dissociations between the conscious awareness and measures of the implicit integration of the spatiotemporal parameters and identities of moving objects. For instance, participants reported seeing a “streaming” visual object while their behavior suggested a (apparently implicit) binding of this object to the feature “bouncing”. These and other observations raise doubts in the idea that feature binding is strongly related to the construction of conscious awareness.

In the present study, we were interested to test whether the consciously perceived coherence or belongingness of two features (operationalized as perceived temporal simultaneity or “occurrence at the same time”) would be systematically related to implicit measures of the binding of the same two features. We varied the temporal relationship between these two features, assuming that people would be less likely to perceive them as belonging to the same event as the temporal interval between them increases. We also assessed the degree of binding between the two features by means of the event file paradigm of Hommel (1998), a variant the object-preview design developed by Kahneman, Treisman, and Gibbs (1992). If conscious awareness would be a direct consequence of feature binding, or even represent the mechanism producing it (as suggested by the global workspace model of Baars, 1988), one would expect that binding would occur only for features that are perceived as belonging to the same event but not for features perceived as belonging to separate events. This was the main hypothesis being tested in Experiment 2 of the present study. The purpose of Experiment 1 was to introduce the multimodal version of the event-file design that we used to assess feature binding in Experiment 2, and to demonstrate that it works with the particular stimuli and parameters chosen.
Experiment 1

An elegant way to test whether people spontaneously bind the codes of the perceptual features of a given event was developed by Kahneman et al. (1992). In a nutshell, these authors presented participants with two visual displays in a row, a task-irrelevant prime display with a number of objects in different locations followed by a probe display with a to-be-identified object. The main finding was that performance was better if the probe object had already appeared in the prime display and, more importantly, that this priming effect was particularly strong if the location of the object was also the same as in the prime display. This observation was taken to suggest that encountering the object in the prime display had led to a binding between object identity and location codes, so that repeating the complete conjunction allowed for a reuse of the same object representation (object file).

Further studies with a stripped-down version of this task revealed that at least part of the effect might not reflect benefits related to the reuse of object representations but, rather, cognitive conflict due to the retrieval of misleading object files. Hommel (1998) presented participants with single-object prime (S1) and probe (S2) displays that repeated or alternated the shape, the color, and/or the location of the stimulus. It turned out that performance was equally good if two or more stimulus features were repeated and if all features were alternated, suggesting that the opportunity to reuse an object file might not provide a particular advantage. However, performance was impaired if one feature was repeated but another alternated, suggesting that the effect reflects interference produced by partial repetitions. If, for instance, participants encounter a red square after having seen a red circle, the repetition of the red color might retrieve the just-created binding of RED and CIRCLE, which creates conflict between the reactivated CIRCLE feature and the actually relevant SQUARE feature (Hommel, 2004).
Recent studies showed comparable interactions between feature-repetition effects for auditory (Mondor, Hurlburt, & Thorne, 2003; Zmigrod & Hommel, 2009) and tactile features (Zmigrod, Spapé, & Hommel, 2009), as well as for combinations of visual and auditory or auditory and tactile features (Zmigrod et al., 2009). This suggests that people spontaneously integrate co-occurring features from various sensory modalities. In the present study, we adopted the design of Zmigrod et al. (2009), which combines visual stimuli varying in color with auditory stimuli varying in pitch (see Figure 6.1). However, given that the design of Experiment 2 required the presentation at different temporal asynchronies, we wondered to which degree feature binding would be affected by temporal asynchrony. Studies on multimodal perception suggest that stimuli that appear within a temporal window of up to about 100 ms (Lewald, Ehrenstein, & Guski, 2001) or even 200 ms (van Wassenhove, Grant, & Poeppel, 2007) are still perceived as being part of the same event. Lewkowicz (1996) has coined this temporal criterion for perceived coherence the “intersensory temporal synchrony window” and has claimed that this window is shorter for frequent multimodal events than unfamiliar or less frequent ones.

Our main question in Experiment 1 was whether at least some degree of asynchrony would be tolerated by the binding process assessed by our task. We thus compared performance in the standard, synchronous version of the task, which amounted to a replication of Zmigrod et al. (2009), with performance in a modified version, where the visual color feature appeared 100 ms after the offset of the auditory feature. As shown in Figure 6.1, each trial started with the presentation of a response cue in the form of a directional arrow, indicating whether a left or right response (R1) was required to the mere onset of S1 (regardless of its features)(FOOTNOTE 1). S2, another audio-visual stimulus, appeared 500 ms after responding to S1. S2 required a binary choice reaction (R2) to the color of the visual feature of S2 (red vs. blue). The two stimulus features varied independently, so that color- and pitch-repetition effects could be analyzed.
Figure 6.1. Experiment 1, overview of the display and the timing of events for the synchronous condition. In the asynchronous condition, the sound preceded the onset of the color by 100 ms.

Method

Participants

Twenty-two participants (4 men) were recruited by advertisement for this experiment and were paid or received a course credit for a 25 min session. Their mean age was 21 years (range 18-30 years). The participants were naïve as to the purpose of the experiment, and they reported not having any sight or hearing problem. They were randomly assigned to two groups, a synchronous (N=11) and an asynchronous (N=11) feature-presentation group.

Apparatus and Stimuli

The experiment was controlled by a Pentium 3 computer attached to a 17-inch CRT monitor. Participants faced the monitor at a distance of about 60 cm
along with headphones. The auditory feature of the stimuli S1 and S2 were composed of two pure tones of 1000Hz and 3000Hz with duration of 50 ms and presented at approximately 70 dB SPL. The visual features of the stimuli S1 and S2 were a blue or a red circle of about 10 cm in diameter. In the synchronous group, the visual and the auditory features were presented at the same time; however, in the asynchronous group, the auditory feature was presented 100 ms before the visual feature. Responses to S1 and to S2 were made by clicking on the left or the right mouse button with index and middle fingers respectively. Response cues were presented in the middle of the screen (see Figure 6.1) with a right or left arrow indicating a left and right mouse click, respectively.

Procedure and Design

The experiment was composed of a practice block with 15 trials and an experimental block with 128 trials. The order of the trials was random. Participants were to carry out two responses per trial: the first response (R1) was a left or right mouse click to the onset of the visual feature of S1 (ignoring its identity) as indicated by the direction of an arrow in the response cue. The second response (R2) was a left or right mouse click to the value of the color dimension of S2. The responses’ mapping was counterbalanced between participants. The participants were instructed to respond as quickly and accurately as possible.

The sequence of events in each trial is shown in Figure 6.1. A response cue with a right or left arrow appeared for 1000 ms to signal R1, which was to be carried out as soon as the color of S1 appeared. The duration between the response cue and S1 was 1000 ms. S2 came up 500 ms after R1, with the color signaling the second response (R2). In the case of incorrect or absent response an error message was presented on the screen.

Results & Discussion

Trials with incorrect R1 responses (0.1%), as well as missing (RT>1200 ms) or anticipatory (RT<100 ms) R2 responses (0.2%) were excluded from
Chapter 6 – Feature Binding and Consciousness

analysis. The mean reaction time for corrected R1 was 222 ms (SD=72.5). From the remaining data, mean reaction times (RTs) and error rates (PEs) for R2 were analyzed as a function of the three variables: the relationship (repetition vs. alternation) between S1 and S2 with regard to color and to pitch, and the presentation type (synchronous vs. asynchronous) (see Table 6.1 for mean RTs and PEs). ANOVAs were performed by using a mixed design with repeated measures on color and pitch repetition and with presentation type as between-group variable.

Table 6.1. Experiment 1: Means of mean reaction times for responses to stimulus 2 (RT R2 in ms) and error rates (in parentheses) as a function of the presentation type (synchronous vs. asynchronous), and the relationship between the stimuli features (S1 – S2) for color and pitch.

<table>
<thead>
<tr>
<th>S1 asynchrony</th>
<th>Color repeated</th>
<th>Color alternated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pitch repeated</td>
<td>Pitch alternated</td>
</tr>
<tr>
<td>0</td>
<td>419 (8.3)</td>
<td>443 (7.5)</td>
</tr>
<tr>
<td>100</td>
<td>417 (6.0)</td>
<td>430 (7.1)</td>
</tr>
</tbody>
</table>

The analysis of the error rates did not reveal any significant effect. For RTs, there was a significant interaction between repetition vs. alternation of pitch and repetition vs. alternation of color, $F(1,20)=22.28$, $p<.0001$. As shown in Table 6.1, responses delayed if one feature was repeated but the other alternated, which is the standard interaction indicative of feature binding (Zmigrod et al., 2009). Importantly, however, this interaction was not modified by presentation type, $p>1$ suggesting binding was unaffected by asynchronous presentation. This was confirmed by separate ANOVAs, which indicated that the color-by-pitch interaction was significant with both synchronous presentation, $F(1,10)=10.42$, $p<.01$, and asynchronous presentation, $F(1,10)=12.54$, $p<.005$. We can thus conclude that visual and auditory stimulus features are spontaneously integrated...
with both perfectly synchronous and slightly asynchronous stimulus onsets, suggesting that intermodal feature integration is using temporarily extended feature-integration windows (Lewald et al., 2001).

**Experiment 2**

Experiment 1 provided evidence that binding effects can be observed with synchronous as well as asynchronous presentation of perceptual features from different modalities, but these findings do not tell us anything about the conscious experience of the participants. Experiment 2 aimed to directly assess this experience by asking participants to judge whether the tone and the color making up S1 would appear at the same time or constitute different events. Given that processing and responding to S2 might affect this experience, we had participants make the judgment online, as a response to the presentation of S1. As Experiment 1 suggests that minor temporal gaps are tolerated by the binding mechanism, we also included larger gaps (of up to 350 ms) in Experiment 2.

One possible problem that might result from introducing larger gaps is that the gap itself might be perceived as coded as a feature of S1. If so, it is conceivable that even strongly asynchronous S1 features are still integrated into the same event representation but this representation is no longer retrieved at S2 processing, simply because the asynchronous S1 and the synchronous S2 are no longer perceived as similar. In other words, the manipulation of synchrony of S1 features may not only affect the likelihood of relating these features to the same event but also introduce a novel feature (i.e., synchrony) that would always mismatch with S2. Without a match, however, the representation of S1 would no longer have the chance to affect S2 processing, and the absence of any effect might be mistaken to imply a lack of S1 feature binding. To test whether this is a real problem, we presented the S2 features (color and pitch) synchronously, as in Experiment 1, or asynchronously (350 ms gap), as in the largest gap condition of
S1. This allowed us to test whether the targeted effects of S1 synchrony would depend on the relationship or similarity between S1 and S2 synchrony.

The synchrony-match issue aside, our main interest was whether spontaneous feature binding, as indicated by the color-by-pitch interactions observed in Experiment 1, would depend on whether the corresponding S1 features would be perceived as belonging to the same event. If so, we would expect a reliable interaction between color- and pitch-repetition effects in trials where participants judge the features as belonging together but no interaction in trials where the two features are perceived as belonging to separate events. The main function of manipulating S1 feature synchrony was to introduce some systematic variability in the judgment, and we expected same-event judgments to become less frequent as the asynchrony increases.

Method

Participants

Twenty participants (1 man) were recruited by advertisement for this experiment and were paid or received a course credit for a 50 min session. Their mean age was 21 years (range 18-30 years), and they fulfilled the same criteria as in Experiment 1.

Apparatus and Stimuli

The apparatus and the stimuli were as in Experiment 1, with the following exceptions. The sound of the first stimulus compound (S1) appeared 0, 50, 150, 250, or 350 ms before the onset of the color. The sound of the second stimulus compound (S2) appeared either 0 or 350 ms before the color. The response cue no longer signaled R1 but contained the judgment-to-key mapping for the response to S1. The participants were instructed to judge whether the sound and the color of S1 appeared “at the same time (together)” or “not at the same time (separately)” and to press the left or right key accordingly.
**Procedure and Design**

The procedure and the design were as in Experiment 1, with the following exceptions (see Figure 6.2). All manipulations were carried out within subjects. There was a practice block of 12 trials and an experimental block with 368 trials. The order of the trials within the blocks was random. Participants were to carry out two responses per trial: a simultaneity judgment of sound and color of S1 (R1) and a left or right response (R2) to the color of S2, as in Experiment 1. The mapping of the stimuli to responses was balanced across participants. In the case of a response omission or an incorrect R2, an error message was presented on the screen.

![Figure 6.2](image_url)  
*Figure 6.2. Overview of the display and the timing of events in Experiment 2.*

**Results & Discussion**

Trials with missing R1 responses (5.5%), as well as missing (RT>1500 ms) or anticipatory (RT<100 ms) R2 responses (0.6%) were excluded from all the analyses. In the following, we will first address the integration effects for all trials regardless of the subjective experience and examine the impact of the S1 and S2 synchrony manipulations, then report the impact of asynchrony on the synchrony judgments, and finally address the impact of this judgment on the integration effects.
Table 6.2. Experiment 2: Means of mean reaction times for responses to stimulus 2 (RT\textsubscript{R2} in ms) and error rates (in parentheses) as a function of S1 asynchrony (in ms) and S2 asynchrony (in ms), and the S1-S2 relationship with respect to color and pitch.

<table>
<thead>
<tr>
<th>S1 asynchrony</th>
<th>S2 asynchrony</th>
<th>Color repeated</th>
<th>Color alternated</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pitch repeated</td>
<td>Pitch alternated</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>512 (4.0)</td>
<td>552 (6.1)</td>
</tr>
<tr>
<td>0</td>
<td>350</td>
<td>457 (9.6)</td>
<td>491 (8.5)</td>
</tr>
<tr>
<td>50</td>
<td>0</td>
<td>505 (6.9)</td>
<td>536 (6.7)</td>
</tr>
<tr>
<td>50</td>
<td>350</td>
<td>465 (6.1)</td>
<td>487 (7.3)</td>
</tr>
<tr>
<td>150</td>
<td>0</td>
<td>528 (8.1)</td>
<td>556 (2.7)</td>
</tr>
<tr>
<td>150</td>
<td>350</td>
<td>457 (6.1)</td>
<td>504 (6.4)</td>
</tr>
<tr>
<td>250</td>
<td>0</td>
<td>541 (6.0)</td>
<td>553 (4.5)</td>
</tr>
<tr>
<td>250</td>
<td>350</td>
<td>483 (5.3)</td>
<td>517 (6.1)</td>
</tr>
<tr>
<td>350</td>
<td>0</td>
<td>531 (5.5)</td>
<td>528 (4.3)</td>
</tr>
<tr>
<td>350</td>
<td>350</td>
<td>476 (5.8)</td>
<td>513 (5.8)</td>
</tr>
</tbody>
</table>

**Binding effects**

Mean RTs and error rates for R2 were analyzed by mean of a four-way ANOVA as a function of S1 asynchrony (0, 50, 150, 250, or 350 ms), S2 asynchrony (0, 350 ms), the relationship between S1 and S2 color (repetition versus alternation), and the relationship S1 and S2 pitch (repetition versus alternation) (see Table 6.2). The analysis of the error rates did not reveal any significant effect. The RTs yielded a significant interaction between color and pitch repetition in RTs, $F(1,19)=20.10$, $p<.0001$, comparable to that obtained by Zmigrod et al. (2009) and in Experiment 1. Importantly this interaction was not modified by S1 asynchrony or S2 asynchrony, $Fs<1$. This shows that multimodal feature binding tolerates temporal asynchronies of at least 350 ms, which implies a rather broad temporal integration window. The only other reliable effect was a
main effect of S2 asynchrony in RTs, $F(1,19)=57.25, p<.0001$, indicating faster performance if the sound preceded the visual presentation by 350 ms (488 ms) than with synchronous presentation (544 ms). Very likely, this observation represents a kind of alerting effect, through which the task-irrelevant tone enhanced the preparation for processing the color stimulus. Importantly, however, there was no indication that the match between S1 and S2 asynchrony would matter.

**Synchrony judgment**

As shown in Figure 6.3, the likelihood of judging the visual and the auditory feature to occur at the same time decreases as the temporal asynchrony increases. This confirms that our manipulation worked as expected.

![Graph](image)

**Figure 6.3.** Percentage of “synchronous” judgments as a function of S1 color-tone asynchrony in Experiment 2.
Impact of conscious experience on binding

We sorted the trials according to the outcome of the synchrony judgment and analyzed RTs by means of a three-way ANOVA with experience (synchronous vs. asynchronous) and pitch and color repetition (vs. alternation) as factors. The only significant result was an interaction between color and pitch, $F(1,19)=25.76$, $p<.0001$, indicative of multimodal binding. This interaction was not modified by subjective experience, $F<1$ (see Figure 6.4). That is, feature integration effects were observed irrespective of whether participants perceived the sound and the color as one or as two different events.

Figure 6.4. R2 reaction times in Experiment 2 as a function of repetition versus alternation of the color and the pitch of S1 and S2, and the perceived simultaneity of color and pitch.
General Discussion

The main question of the present study was whether intermodal feature binding and the conscious perception of multi-modal features as belonging to the same event are related. If they would, the probability of feature binding should have been correlated with the probability of perceiving the bound features as belonging to the same perceptual event or, more specifically, as occurring at the same time. And yet, our findings do not provide any evidence for such a relationship. In fact, binding effects were entirely unrelated to conscious perception and did not even decrease in size when the bound features were perceived as separate events.

This observation fits with previous reports of dissociations between binding effects and conscious perception (Mitroff et al., 2005; Wojciulik & Kanwisher, 1998) and challenges, or at least helps to refine theoretical accounts that claim or suggest a tight relationship between binding and consciousness. On the one hand, one can argue that the fact that binding is not necessarily reflected in conscious perception is not inconsistent with the assumption that feature integration is a necessary precondition for coherent conscious perception (Crick & Koch, 1990; Engel & Singer, 2001; Treisman, 2003). It is possible that binding is a necessary first step which however needs to be followed up by other processes to generate a conscious impression (LaRock, 2007; van Leeuwen, 2007). In our study, binding might have taken place while these other processes did not, so that we were able to measure binding after effects independent of conscious experience. However, not only would such an approach beg the question of what these other processes might be and why they failed to take place in the present study, but we would also need to explain why participants were able to make synchrony judgments that apparently reflected their conscious experience. If a conscious representation was constructed while binding took place, why was the outcome of binding not reflected in the conscious
representation? Even more difficult to apply to our findings is the idea that integration across specialized modules requires a processing state that is correlated with conscious awareness (Baars, 1988; Dehaene & Naccache, 2001). If integration is impossible without such a conscious state, how is it possible that the outcome of binding processes is not reflected in conscious awareness?

These considerations suggest that binding seems to operate independently of conscious awareness, which again implies that it solves processing problems other than the construction of conscious representations. As pointed out already, our ability to carry out multiple actions at (about) the same time requires some sort of feature integration, so that concurrently active action routines “know” which objects they are to process. Given the evidence that conscious awareness does not seem to play an important role in online-controlling such actions (Hommel, 2000, 2007a), it makes a lot of sense that feature integration operate independently of consciousness. Moreover, various authors have claimed that the human brain is proactive and constantly generating unconscious predictions about upcoming events (Bar, 2009; Neisser, 1976; Schubotz, 2007; Zacks, Speer, Swallow, Braver & Reynolds, 2007). These kinds of predictions must rely on memory traces that integrate the features belonging to the same event, suggesting that they require feature binding as well. Hence, the processes responsible for constructing conscious representations are by no means the only possible clients will feature-binding operations and, as our findings suggest, they may not even be the most important ones.
Footnote

1. Having participants respond to S1 (as in the standard setup of Hommel, 1998) was not a strict requirement for the logic of Experiment 1. However, given that we needed a response to S1 in Experiment 2, including such a response in Experiment 1 already made the two experiments more similar and, thus, easier to compare. Moreover, previous studies have shown that people do not only bind stimulus features but stimulus and response features as well (Hommel, 1998, 2004), so that we were interested to look into these effects for explorative purposes. However, given that, in addition to the standard stimulus-response-interaction effects reported earlier (Zmigrod et al., 2009), no interaction with the synchrony manipulation was obtained, we do not present response-related effects for the sake of clarity.
Chapter 7

General Discussion and Conclusions
Chapter 7 – General Discussion

General Discussion

Our brain is continuously bombarded with many types of information delivered by various sensory channels and processed in distinct cortical regions (DeYoe & van Essen, 1988; Wessinger et al., 2001). How the brain integrates this distributed information into a coherent representation (the so-called binding problem) is the topic of the present thesis. In particular, we sought to explore principles and constraints of integration processes within and between different sensory modalities and action planning. The following findings emerged from this endeavor:

Principles underlying multimodal integration

First, as mentioned earlier, feature integration has been explored mostly in the visual domain (Allport, Tipper, & Chmiel, 1985; Treisman & Gelade, 1980: object file), and across visual and action planning (Hommel, 1998; 2004: event files). Here, we extended these findings to other modalities and across multimodal perception and action, substantiating the event files concept. In chapter 2, we demonstrated feature integration in the auditory domain and across auditory and action planning. Further, in chapters 3 and 4, we expanded these findings across different domains such as vision, audition, taction and action planning. Hence, we provided empirical evidence that feature integration crosses borders between modalities and domains, supporting the idea that it is a general phenomenon in perception and action, as suggested by Hommel (2004).

Second, feature integration seems to occur rather automatically, and does not require focused attention. In chapter 2, we presented evidence that auditory features (such as pitch and loudness or pitch and location) were integrated together while only one of the features was task relevant and the other could be ignored, similarly to the findings in the visual domain (Hommel, 1998; Hommel & Colzato, 2004). In addition, in chapters 3 and 4, we demonstrated that spontaneous integration is not limited to a single modality, but rather occurs
between features from different modalities such as visual, auditory, and tactile. Thus, feature integration does not seem to require attention (Hommel, 2005); yet not all the features of an event get integrated, suggesting that this process is not fully automatic.

Third, even though attention is not a strict necessity for feature binding, it plays a prominent role in modulating the integration processes. Integration effects, evidently, are more pronounced for the task relevant features. Specifically, features that are related to the task, thus being attended, are more likely to be integrated than unrelated features. For instance, in chapter 2, we observed higher partial repetition cost (an index for the strength of the binding) between response and the task relevant feature (pitch in the pitch task and loudness in the loudness task). Moreover, this pattern was even more apparent in chapters 3 and 4, as the irrelevant feature came from a different modality. Here, the stimulus-response integration was clearly influenced by which sensory modality was task-relevant, indicating that features from the same domain are more likely to be integrated. Furthermore, features that vary on the task-relevant dimensions are more likely to become part of the binding than features unrelated to these dimensions. For instance, in chapter 2, when we used non-spatial response, the irrelevant location feature was less likely to be bound than when we employed spatial response. Hence, attention has a significant and pronounced function in feature integration.

Fourth, apart from attention, there are other factors affecting the likelihood of a feature to be integrated and/or retrieved. In chapter 3, we found that salience of a feature affects this probability. For instance, the auditory feature pitch was less sensitive to task relevance (either color or vibration) and was more likely to be integrated with the response, similar to the findings of Dutzi & Hommel (2009). Another interesting observation concerns the feature’s modality. As described in chapter 3, the audio-tactile interaction was numerically larger in the tactile task than in the auditory task, and the visio-audio interaction was more
pronounced in the auditory task than in the visual task. Apparently, this pattern tactile $\rightarrow$ auditory $\rightarrow$ visual, also followed the ontogenetic development of cortical multisensory integration as well as the emergence of the multisensory neurons (Wallace et al., 2006). Thus presumably, this ontogenetic development affects the strength of the connections between sensory modalities.

Fifth, another central principle which was introduced in chapter 2 is the temporal-overlap principle. This principle posits that features of events (perceptual or response features) are bound together if the activations of their codes overlap in time. This principle encompasses all other factors affecting the intensity or the duration of feature code’s activation, in such a way that the likelihood of a feature to be integrated increases as a function of the strength and the endurance of its activation. For instance, properties such as task relevance or salience are likely to induce stronger activation of the feature’s code (Found & Müller, 1996; Hommel, Müsseler, Aschersleben, & Prinz, 2001) and are thus, according to this principle, more likely to be involved in the integration. In chapter 2, we presented a model explaining the probability of auditory features and response features to be integrated, and suggested that salient features (as loudness was in this setting) are more probable to be part of the integration. Additionally, in chapter 4, we observed less pronounced effects in a multimodal context than unimodal, suggesting a weaker temporal overlap of features coming from other modalities than the same modality. Hence, the integration of event features (stimulus and/or response) is sensitive to temporal characteristics.

Sixth, the endurance characteristic of the bindings also plays a central role in the lifetime of the episodic memory traces. Studies showed that the feature binding in the visual domain (object file) can last at least 8 sec (Noles, Scholl, & Mitroff, 2005) and feature integration across visual perception and action (event file) can last at least 4 sec (Hommel & Colzato, 2004). In chapter 4 we presented findings indicating similar persistence patterns between auditory bindings and visual bindings (Hommel & Colzato, 2004) in terms of partial repetition costs and
decay rate. Additionally, the multimodal bindings, although exhibiting lower partial repetition costs (weaker strength), showed a comparable decay rate as the unimodal bindings. With regard to the stimulus-response bindings, the findings show similar decay rates for the relevant and the irrelevant stimulus both in the unimodal and multimodal context, suggesting a spontaneous decrease of activation over time. However, the strength of the activation (the function’s intercept) was more pronounced for the relevant feature and response binding than the irrelevant one. Hence, we concluded that the lifetime of a specific feature-binding depends on the initial strength of activation of its features’ code.

Seventh, focusing on behavioral studies, one way to look at the dynamics of the binding mechanism is by delving into the essence of partial repetition costs, which reflects changes in the event representation. Findings show that a complete repetition of both features (e.g. in chapter 4: S1=high pitch+red $\rightarrow$ S2=high pitch+red) or a complete alternation (e.g. in chapter 4: S1=high pitch+red $\rightarrow$ S2=low pitch+blue) produces better performance than do partial repetitions (e.g. in chapter 4: S1=high pitch+red $\rightarrow$ S2=high pitch+blue, or S1=high pitch+red $\rightarrow$ S2=low pitch+red). These partial repetition costs suggest that repeating at least one feature either from the stimulus or the response leads to the retrieval of the previously created episodic traces, which interferes with current processing if that involves the reactivation of a no longer valid feature code (Hommel, 2004). The likelihood of a successful retrieval of the previous traces depends on the relevance of the dimensions to the task (chapter 2), the salience of the features (chapter 3), the spontaneous decay of the bindings (chapter 4) and the width of the temporal window between the events (chapter 4). Hence, feature binding is a highly dynamic process which a number of factors mediate its lifetime.

In addition, to ensure efficiency and accuracy of the new event representation, the process of handling and managing these previous traces seems to require some sort of control. Thereby, well balanced control process can accelerate and promote efficiency of the binding processes, and can maintain
rebinding costs at an optimal level. In chapter 5, we compared the partial repetition costs between typically developing children and children with a neurodevelopmental disorder such as autistic spectrum disorder (ASD) who exhibit a deficit in control and flexibility as measured by executive functions (for reviews, see Corbett, Constantine, Hendren, Rocke, & Ozonoff, 2009; Hill, 2004). The results revealed higher stimulus-response partial repetition costs for the ASD children than the typically developing children, suggesting an inefficient handling of event representation by ASD. These costs were also correlated with control and flexibility indices as measured by an executive function task (set-shifting task in this study). These findings are also consistent with higher partial-repetition costs found in individuals with low fluid intelligence (Colzato, van Wouwe, Lavender, & Hommel, 2006), and in young children and elderly participants, as compared to young adults (Hommel, Kray, & Lindenberger, submitted). Thus, these observations provide insight into the role of control processes in the management of episodic feature bindings and a link to the dopaminergic system, presumably modulating these processes (Colzato, van Wouwe, & Hommel, 2007a; Colzato, Kool, & Hommel, 2008; Colzato & Hommel, 2008).

Last, many definitions of the binding problem included the notion of unified/coherent perception and even subjective awareness (van de Velde, 2007), thus associating consciousness with binding. However, these definitions are now in dispute in the consciousness literature (Revonsuo, 1999), and yet not many empirical foundations exist. When considering unified perception and binding one can wonder whether unified perception is a necessity for integration to occur, or whether unified perception is a cause or a consequence of bindings. In chapter 6, we provided empirical evidence that this is not the case, by employing different temporal onset on the event’s features, and linking the integration effects with the subjective experience of the participants. The findings show that unified perception is more sensitive to temporal synchronization than integration effects, thus dissociating between the two processes. Therefore, there is no direct
correspondence between unified perception and feature integration, and unified perception is neither prerequisite nor a consequence of binding.

In sum, the process of feature integration across modalities and domains does not seem to be a unitary process but rather to involve different assemblies of local and global networks, which operate within modalities and between modalities and domains (see Figure 7.1). Indeed, Keetels, Stekelenburg, & Vroomen (2007) demonstrated that auditory grouping of a sound preceded multisensory integration. This suggests constructions of local networks inside a modality, such as the visual or the auditory modality, and only later constructions of more global networks across other modalities and domains in the form of event files. Furthermore, various studies showed that different neuromodulatory systems are mediating binding within and across modalities. For example, the perceptual binding (such as visual binding) seems to be modulated by the muscarinic–cholinergic system (Colzato, Fagioli, Erasmus, & Hommel, 2005), and the dopaminergic system seems to be associated with integration across perception and action (Colzato, van Wouwe, & Hommel, 2007a). In chapter 4, we provided evidence that each of the various integration effects (within modality, across modalities, between perception and action) has different activation strength (in terms of partial repetition costs) which may indicate separate networks. This pattern is also reflected, as delineated in chapter 4, in the lifetime of a specific feature binding, a longer activation within modalities than across modalities. Thus, it appears that the event integration is not a unitary process but rather a graded process composed of various assemblies, each can decay at a different pace, based on their initial strength and onset.
Figure 7.1. A schematic model of multimodal feature integration. The figure shows dynamic associations of event representations (network rings) as a set of interconnected unimodal representations (hubs). Different levels of event files have been left out for clarity. Maps of visual features are drawn in the occipital lobe, auditory maps are drawn in the temporal lobes, tactile maps are drawn in the parietal lobe, and action codes are drawn within the frontal lobe. The connections are reciprocal, and its width indicates its strength of activation. The straight lines represent an activation of task-relevant dimensions while the dotted lines represent connections to task-irrelevant features. Unimodal features are integrated in local networks, which later interconnect to other modalities and domains.
Conclusions

The aim of the current thesis was to extend our understanding of how distributed feature codes (from same and different domains) lead to a unified and coherent representation. The empirical evidence presented in this thesis suggests that feature integration emerges in a graded manner through intra and inter connections within and across modalities and domains, employing general principles (such as temporal overlap of feature activation), which capture the role of attention, salience and dominance among the features and the domains. Also, we argued that control processes are important in handling and maintaining retrieved traces for efficient use. Nevertheless, the binding mechanism does not require a conscious or unified perception and unified perception is not the outcome of feature binding. Thus, what is the purpose of the binding mechanism, if not for consciousness, is a matter of a further philosophical inquiry and extensive empirical research.
References


156
References


References


Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus-response translation. In S. Monsell, & J. Driver (Eds.), *Control of


References


References


References


Samenvatting

Een van de meest opmerkelijke aspecten van multimodale perceptie is haar intrinsieke samenhang: hoewel informatie via verschillende kanalen wordt verwerkt in diverse hersengebieden, vormt onze bewuste ervaring toch telkens weer een eenheid. Bij het eten van een sandwich bijvoorbeeld, is de informatie-integratie van verschillende dimensies belangrijk: de visuele eigenschappen zoals de kleur en vorm van de ingrediënten, de tactiele eigenschappen zoals de structuur en de temperatuur van de sandwich, de chemische eigenschappen zoals geur en smaak, en dat alles gecombineerd met de actie van het kauwen en de eventueel daardoor geproduceerde geluiden. Hoe het brein al deze verschillende typen van informatie, verwerkt in verschillende hersengebieden, samenvoegt tot een geïntegreerde eenheid wordt het ‘binding probleem’ genoemd (Treisman, 1996).


Deze
mate van afname is vergelijkbaar voor unimodale en multimodale bindings, wat wijst op een gedeeld onderliggend mechanisme. In hoofdstuk 5 wordt onderzocht hoe gedragssturing en flexibiliteit geassocieerd zijn met ‘event file maintenance’. Hierbij worden integratie-effecten tussen populaties met ontwikkelingsstoornissen zoals kinderen met een autisme spectrum stoornis (ASS) en normaal ontwikkelende kinderen vergeleken. Deze studie toont aan dat ASS kinderen beperkingen hebben in het bijwerken van ‘event file’ representaties vanwege een gebrek aan cognitieve flexibiliteit zoals gemeten in een executieve functie taak. Dit suggereert een overeenstemming tussen deze cognitieve functies die vermoedelijk veroorzaakt wordt door prefontale dopaminerge hypoactiviteit. In het laatste empirische hoofdstuk 6 wordt de relatie tussen ‘feature binding’ en coherente perceptie (ons bewustzijn) onderzocht waarbij wordt gesteld dat binding processen noodzakelijke voorwaarde noch gevolg zijn van perceptie die een eenheid vormt. Met deze studie wordt een door sommigen veronderstelde symbiotische relatie tussen deze twee processen verbroken en worden vraagtekens geplaatst bij bepaalde gebruikelijke definities in de literatuur. Tenslotte worden in hoofdstuk 7 de principes en beperkingen van ‘feature integration’ in een schematisch model gepresenteerd.

Samenvattend kan op basis van dit onderzoek gesteld worden dat ‘feature integration’ over modaliteiten en domeinen heen niet een enkelvoudig en algemeen proces is, maar eerder verschillende samenstellingen van lokale en globale neurale netwerken veronderstelt die binnen modaliteiten en tussen modaliteiten en domeinen een rol spelen.
Acknowledgements

The last 5 years in Leiden were of the most intriguing in my life. Not only have I followed my dream studying the brain, but I also grew as a researcher and a thinker. This would not be possible without some of the amazing people whom I met in Leiden. First, I am heartily thankful to my supervisor, Bernhard Hommel, whose guidance and support enabled me to develop an understanding of the subject and the scientific inquiry, all the comments, suggestions throughout the years and his academic experience were invaluable to me.

I am indebted to Lorenza whose informal support and encouragement as well as her collaboration and academic advices helped me greatly with my research. I also wish to thank my best friend and roommate in Leiden Szilvia for helping me get through the difficult and the happy times, and for the academic advises, the emotional support, friendship, fun, and care. I am grateful to Albertien for providing not only administrative and professional help throughout the years but also for kindness and a caring environment.

Last but not least, I wish to thank to my family, my parents, Mira and Jacob, who always believed in me, and helped me to become who I am; my sister Orit and my brother Avishai for supporting and caring. My husband Shai whose enormous support, encouragement, and love are the real force behind my studies, and my wonderful children, Leor, Ran and Shani who are my real inspiration and the reason for everything in my life. Words cannot express how privileged I am for having such amazing and supporting family. To them, I dedicate with love this thesis.
Curriculum Vitae

Sharon Zmigrod was born in Israel on April 10, 1967. In 1985, she completed secondary education at the Head Ham high school in Israel. In 1989, she started studying psychology and computer science in Bar-Ilan University. In 1990, she moved to University of New Haven, where she completed B.Sc. in computer science with distinction. After graduating in computer science, she worked in the Hi-Tec industry, for several years, in various projects while completing MBA program from Edinburgh University. In 2002, she switched back to psychology, and graduated with distinction in Experimental Psychology from Bar–Ilan University. Since 2004, she has worked as a PhD student at Leiden University on feature integration across multimodal perception and action. In addition to the scientific work, she involves in a number of educational tasks in the International School of Amsterdam where her 3 children are studying.