

# Temporal dynamics of unimodal and multimodal feature binding

SHARON ZMIGROD AND BERNHARD HOMMEL

Leiden University, Leiden, The Netherlands  
and Leiden Institute for Brain and Cognition, Leiden, The Netherlands

In two experiments, we 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 when the second of two consecutive stimuli shared all or none of the features of the first, rather than when only one of the features overlapped. Comparable partial-overlap costs were obtained for combinations of stimulus features and responses. These effects decreased systematically with increasing time between the two stimulus-and-response events, and the decreased 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.

One of the challenges human perception poses is understanding how the brain binds codes of features within and across sensory modalities, despite these codes’ being processed in various cortical areas (e.g., Goldstein, 2007; Wessinger et al., 2001; Zeki & Bartels, 1999). This so-called *binding problem* was investigated initially in the visual domain (see, e.g., Allport, Tipper, & Chmiel, 1985; Kahneman, Treisman, & Gibbs, 1992; Treisman & Gelade, 1980), then in the auditory domain (see, e.g., Dyson & Quinlan, 2004; Hall, Pastore, Acker, & Huang, 2000; Takegata et al., 2005), and more 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 sequential-effects studies, participants typically carry out two responses in a row (see Figure 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 stimulus onset asynchrony (SOA) or response–stimulus interval (RSI), 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

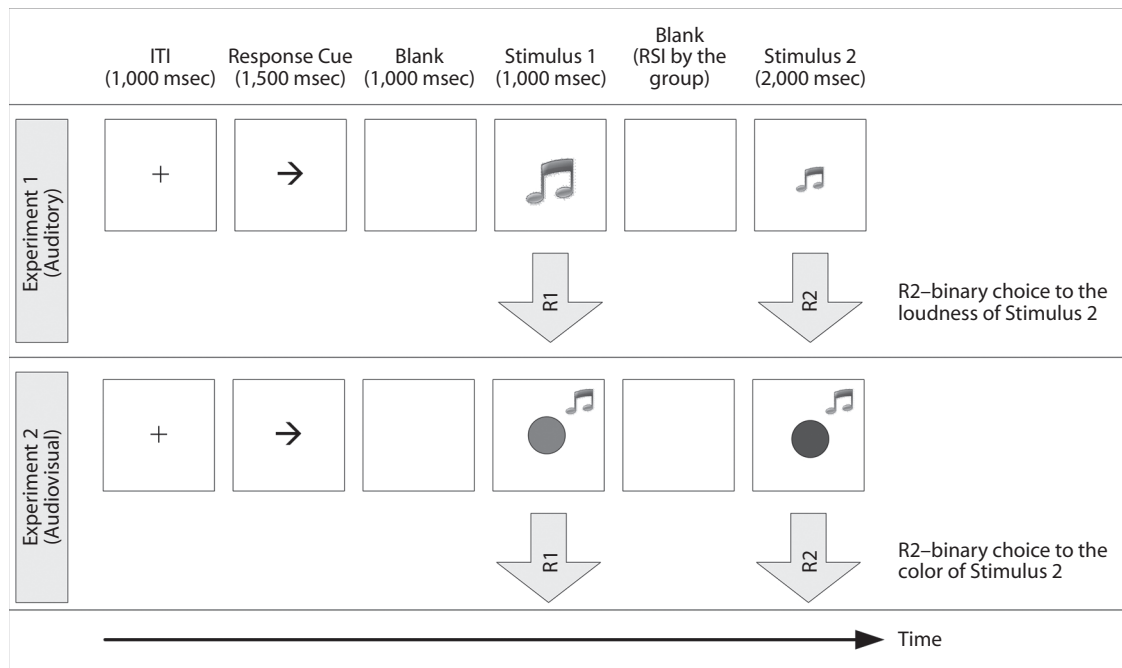
of stimulus-feature repetition were obtained. Yet more interesting, interactions between different stimulus-feature repetition effects and between stimulus-repetition and response-repetition effects were observed for visual features and response (Hommel, 1998, 2005), auditory features and response (Mondor, 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, that S1 and S2 consist of varying combinations of auditory pitch (low vs. high) and visual color (red vs. blue), as in Zmigrod et al.’s (2009) study. Findings show that a complete repetition of both features (e.g., S1 = blue + low → S2 = blue + low) or a complete alternation (e.g., red + high → blue + low) produces better performance than do partial repetitions (e.g., 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 the coding of 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 the coding

---

S. Zmigrod, szmigrod@fsw.leidenuniv.nl

---



**Figure 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). Depending on the group, the second stimulus, S2, appeared 500, 1,000, 2,000, 3,000, or 4,000 msec after R1. S2 signaled R2, a speeded left or right mouse button click, according to the instructed mapping and task.

of *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-ignored flankers were correlated with responses—that stimulus–response bindings can be created without attention. In the auditory domain, Takegata and colleagues (2005) showed that feature integration can occur without focused attention on the to-be-integrated features. In Takegata et al.'s 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 and 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 have demonstrated that irrelevant features from the same modality as the attended feature (in vision, see Hommel, 2005; in audition, see Dyson & Quinlan, 2004) or from a nonattended 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 do 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 past two or three decades, and many studies have investigated the interplay among 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 Gelder, 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 audiovisual 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 prevent the use of tight temporal-synchrony criteria for crossmodal binding). And yet, our conscious perception of multimodal stimuli is commonly coherent and unified, suggesting that binding works. We tested different properties to find the glue that 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, in press; Mitroff & Alvarez, 2007). Yet, no definitive 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 that suggests 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 a 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 we 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 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 (e.g., Hommel & Colzato, 2004) have provided evidence for both relative stability (i.e., the relationship between the two events matters, even with intervals of 4 sec) and decay<sup>1</sup> (i.e., effect sizes tend to decrease over time). Considering these observations, our present study of unimodal auditory binding (Experiment 1) and of multimodal audiovisual binding (Experiment 2) was aimed at addressing three questions of theoretical relevance.

First, we wanted 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. We did this 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. 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 us to test 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) in both the unimodal experiment with auditory stimuli (Experiment 1) and the multimodal experiment with audiovisual 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 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 500, 1,000, 2,000, 3,000, or 4,000 msec after the response to S1, depending on the RSI (which varied between participants). The response to S2 (R2) was a binary-choice reaction according to the task-relevant feature of S2. In Experiment 1, the feature was loudness (loud vs. soft); in Experiment 2, the feature was color (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 important, their interactions (taken to indicate bindings) could be studied. One of the stimulus features was task relevant because it signaled R2 (loudness in Experiment 1 and color in Experiment 2), and one was task irrelevant (pitch, in both experiments); 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 inter-

act, 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

In Experiment 1, we used unimodal stimuli and manual responses, with loudness being the relevant stimulus feature and pitch being 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 4,000 msec (amounting to SOAs of about 800–4,300 msec), similar to Hommel and Colzato's (2004) unimodal visual study.

### Method

**Participants.** Fifty-five Leiden University students (7 male) ranging in age from 18 to 27 years ( $M = 20$  years) were recruited by advertisement and were paid or received course credit for a 20-min session. All of the participants reported having no known sight or hearing problems. The participants were naive as to the purpose of the experiment. The participants were randomly but equally assigned to five groups with different RSIs (500, 1,000, 2,000, 3,000, and 4,000 msec).

**Apparatus and Stimuli.** The experiment was controlled by a Targa Pentium 3 based computer attached to a Targa TM 1769-A 17-in. monitor. Participants faced the monitor at a distance of about 60 cm and wore headphones. The experiment's stimuli S1 and S2 were two pure tones of 1,000 and 3,000 Hz with durations of 50 msec and were presented at 60 and 80 dB SPL, respectively. Response cues were presented in the middle of the screen (see Figure 1), with a right- or left-facing arrow indicating a left or right mouse click, respectively. Responses to S1 (serving as a mere "go" signal) and to the loudness of S2 were made by clicking on the left or right mouse button, respectively, with the same hand.

**Procedure and Design.** The experiment comprised a practice block with 15 trials and an experimental block with 96 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; the other half of the participants 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 1 (upper panel). A response cue with a right- or left-facing arrow was visually presented for 1,500 msec, signaling response (R1), which was to be carried out after S1 was presented. S2 appeared 500, 1,000, 2,000, 3,000, or 4,000 msec after the onset of R1 (i.e., the response to S1). In case of an incorrect or absent response, an error message was presented.

### Results and Discussion

Responses with mean reaction times (RTs) greater than 1,200 msec were considered *missing*; those with mean RTs less than 100 msec were considered *anticipatory*. Trials with incorrect R1 responses (0.3%), as well as those with missing or anticipatory R2 responses (0.05%), were excluded from analysis. The mean RT for correct R1 was 284 msec ( $SD = 80$ ). From the remaining data, mean RTs and percentages of errors (PEs) for R2 were analyzed as a

function of the four variables: (1) the relationship between S1 and S2 (repetition vs. alternation), with regard to loudness; (2) the relationship between S1 and S2 (repetition vs. alternation), with regard to pitch; (3) the relationship between R1 and R2 (repetition vs. alternation); and (4) the RSI condition (500, 1,000, 2,000, 3,000, or 4,000 msec). ANOVAs were performed by using a mixed design with repeated measures on three variables and with RSI as the between-groups 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 msec and 9.4%, respectively) than in alternating trials (532 msec 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 quicker responses in repetition trials (532 msec) than in alternation trials (552 msec). Replicating earlier findings (see Zmigrod & Hommel, 2009), the results further revealed interactions between the relevant and irrelevant stimulus features 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). We grouped our observations according to these theoretical implications.

First, there were a number of effects involving the repetition of loudness and pitch, the relevant and irrelevant stimulus features, respectively (for means, see Table 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 impaired performance when the other auditory feature was alternated, not 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 when 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 the 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 1**  
Experiment 1: Means of Mean Reaction Times (in Milliseconds) for Responses to Stimulus 2 ( $RT_{R2}$ ) As a Function of the Response–Stimulus Interval (RSI) and the Relationship Between the Stimuli Features (S1–S2) for Loudness and Pitch

RSI	Loudness Repeated		Loudness Alternated		Partial Repetition Cost
	Pitch Repeated	Pitch Alternated	Pitch Repeated	Pitch Alternated	
500	565	645	589	570	49
1,000	516	563	547	514	40
2,000	476	538	479	467	37
3,000	516	577	526	531	28
4,000	554	578	554	537	20

**Table 2A**  
**Experiment 1: Means of Mean Reaction Times (in Milliseconds) for Responses to Stimulus 2 (RT<sub>R2</sub>), As a Function of the Response–Stimulus Interval (RSI), the Relationship Between the Responses (R1 and R2), and the Relationship Between the Stimuli Features (S1 and S2) for Loudness**

RSI	Response Repeated		Response Alternated		Partial Repetition Cost
	Loudness Repeated	Loudness Alternated	Loudness Repeated	Loudness Alternated	
500	560	597	650	562	63
1,000	514	552	565	510	46
2,000	489	484	525	462	29
3,000	531	543	563	514	30
4,000	552	556	581	536	24

**Table 2B**  
**Experiment 1: Means of Mean Reaction Times (in Milliseconds) for Responses to Stimulus 2 (RT<sub>R2</sub>), As a Function of the Response–Stimulus Interval (RSI), the Relationship Between the Responses (R1 and R2), and the Relationship Between the Stimuli Features (S1 and S2) for Pitch**

RSI	Response Repeated		Response Alternated		Partial Repetition Cost
	Pitch Repeated	Pitch Alternated	Pitch Repeated	Pitch Alternated	
500	542	616	613	599	44
1,000	510	556	554	521	39
2,000	462	511	493	495	23
3,000	511	563	532	545	19
4,000	545	563	564	553	14

Second, the effects of loudness (the relevant stimulus feature) and response repetition (for means, see Table 2A) interacted in RTs [ $F(1,50) = 60.91, p < .001$ ] and PEs [ $F(1,50) = 21.25, p < .001$ ]. Performance was impaired when loudness was repeated but the response was 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 2B) interacted in RTs [ $F(1,50) = 23.66, p < .001$ ] and PEs [ $F(1,50) = 39.81, p < .001$ ], due to worse performance when pitch was repeated but the response was 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 [for 500 msec,  $F(1,10) = 26.26, p < .001$ ; for 1,000 msec,  $F(1,10) = 6.74, p < .05$ ; for 2,000 msec,  $F(1,10) = 5.43, p < .05$ ; for 3,000 msec,  $F(1,10) = 13.5, p < .005$ ; for 4,000 msec,  $F < 1$ ].

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 ( $RT_{\text{partial repetition}} - RT_{\text{complete repetition/alternation}}$ ) per effect type<sup>2</sup> and RSI and ran an ANOVA with RSI and effect type (loudness–pitch,

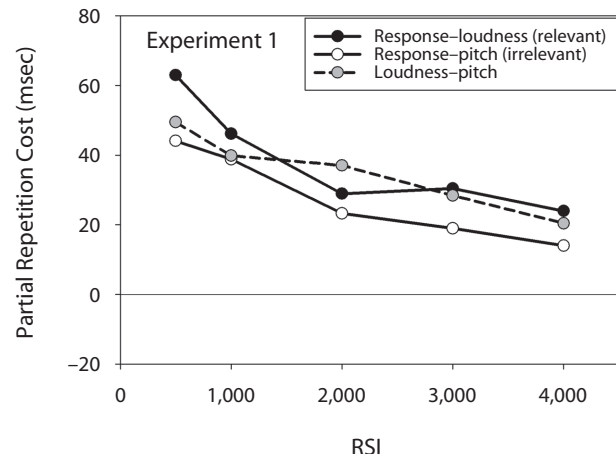
loudness–response, and pitch–response) as factors. The only result this produced was a significant main effect of RSI [ $F(4,50) = 3.01, p < .05$ ]. We take these observations to suggest that the costs produced by partial repetition decrease over time for all types of effects, and they do so in rather comparable ways (Figure 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 when the relevant stimulus feature is involved, which is consistent with previous hints toward 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 2

Experiment 1 provided evidence that bindings involving auditory features are relatively robust, but there was also substantial decay over time. We conducted Experiment 2 to see whether a comparable pattern could be found for intermodal integration. Accordingly, we replaced the auditory feature of loudness with the visual feature of color as the task-relevant variable (i.e., as the variable that signaled R2).

### Method

**Participants.** Fifty-five Leiden University students (5 male) ranging in age from 18 to 30 years ( $M = 20.5$  years) participated for a 20-min session. All of the participants reported having no known sight or hearing problems. The participants were naive as to the purpose of the experiment. The participants were randomly assigned to five groups with different RSIs (500, 1,000, 2,000, 3,000, and 4,000 msec).



**Figure 2.** Partial repetition costs, as a function of response–stimulus interval (RSI), in Experiment 1.

**Table 3**  
**Experiment 2: Means of Mean Reaction Times (in Milliseconds)**  
**for Responses to Stimulus 2 (RT<sub>R2</sub>), As a Function of the**  
**Response–Stimulus Interval (RSI) and the Relationship**  
**Between the Stimuli Features (S1–S2) for Color and Pitch**

RSI	Color Repeated		Color Alternated		Partial Repetition Cost
	Pitch Repeated	Pitch Alternated	Pitch Repeated	Pitch Alternated	
500	438	456	465	458	13
1,000	467	478	457	457	5
2,000	421	441	410	420	5
3,000	514	519	511	521	–2
4,000	494	495	468	484	–8

**Apparatus and Stimuli.** The apparatus was the same as in Experiment 1. The auditory stimuli were pure tones of 1,000 and 3,000 Hz, with durations of 50 msec, presented at approximately 70 dB SPL, and they were accompanied by a blue or red circle at the center of the monitor.

**Procedure and Design.** We used the same setup as in Experiment 1, but 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 comprised a practice block with 15 trials and an experimental block with 128 trials. The sequence of events in each trial is shown in Figure 1 (lower panel).

## Results and Discussion

Trials with incorrect R1 responses (1%), as well as those with missing (RT > 1,200 msec) or anticipatory (RT < 100 msec) R2 responses (0.05%), were excluded from analysis. The mean RT for correct R1 was 239 msec ( $SD = 75$ ). Analogous to Experiment 1, mean RTs and PEs for R2 were analyzed as a function of the four variables: (1) the relationship between S1 and S2 (repetition vs. alternation), with regard to color; (2) the relationship between S1 and S2 (repetition vs. alternation), with regard to pitch; (3) the relationship between responses R1 and R2 (repetition vs. alternation); and (4) the RSI condition (500, 1,000, 2,000, 3,000, or 4,000 msec).

Similar to the findings of 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 msec) than when it was alternated (475 msec). 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 interactions were of greater theoretical interest.

First, we analyzed the effects involving color and pitch, the relevant and the irrelevant stimulus feature, respectively (for means, see Table 3). There was no 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 msec) [ $F(1,10) = 6.51$ ,  $p < .05$ ]. This result conceptually replicates the finding of Zmigrod et al. (2009), who also used a rather short interval between the events. However, in comparison with 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.

Second, the effects of color and response repetition (for means, see Table 4A) interacted in RTs [ $F(1,50) = 67.04$ ,  $p < .001$ ] and PEs [ $F(1,50) = 49.32$ ,  $p < .001$ ]. Performance was impaired when 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 RSI [for 500 msec,  $F(1,10) = 10.54$ ,  $p < .01$ ; for 1,000 msec,  $F(1,10) = 23.00$ ,  $p < .001$ ; for 2,000 msec,  $F(1,10) = 6.52$ ,  $p < .05$ ; for 3,000 msec,  $F(1,10) = 18.71$ ,  $p < .01$ ; for 4,000 msec,  $F < 1$ ].

Third, the effects of pitch and response repetition (for means, see Table 4B) interacted in RTs [ $F(1,50) = 4.51$ ,  $p < .05$ ] and PEs [ $F(1,50) = 19.26$ ,  $p < .001$ ], indicating worse performance with partial repetitions of either pitch or response, as compared with 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-msec [ $F(1,10) = 10.25$ ,  $p < .01$ ] and 1,000-msec [ $F(1,10) = 19.39$ ,  $p < .001$ ] RSIs, 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.

To analyze the temporal dynamics of bindings, we again calculated partial repetition costs (RT<sub>partial repetition</sub> – RT<sub>complete repetition/alternation</sub>) 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 effect type [ $F(2,100) = 22.12$ ,

**Table 4A**  
**Experiment 2: Means of Mean Reaction Times (in Milliseconds)**  
**for Responses to Stimulus 2 (RT<sub>R2</sub>), As a Function of the**  
**Response–Stimulus Interval (RSI), the Relationship Between**  
**the Responses (R1 and R2), and the Relationship Between**  
**the Stimuli Features (S1 and S2) for Color**

RSI	Response Repeated		Response Alternated		Partial Repetition Cost
	Color Repeated	Color Alternated	Color Repeated	Color Alternated	
500	425	495	469	428	56
1,000	445	475	501	440	45
2,000	416	429	446	401	29
3,000	501	535	532	498	34
4,000	480	496	509	456	35

**Table 4B**  
**Experiment 2: Means of Mean Reaction Times (in Milliseconds)**  
**for Responses to Stimulus 2 (RT<sub>R2</sub>), As a Function of the**  
**Response–Stimulus Interval (RSI), the Relationship Between**  
**the Responses (R1 and R2), and the Relationship Between**  
**the Stimuli Features (S1 and S2) for Pitch**

RSI	Response Repeated		Response Alternated		Partial Repetition Cost
	Pitch Repeated	Pitch Alternated	Pitch Repeated	Pitch Alternated	
500	447	472	455	441	20
1,000	455	464	469	471	3
2,000	411	434	420	427	8
3,000	511	525	515	516	6
4,000	480	496	482	483	8

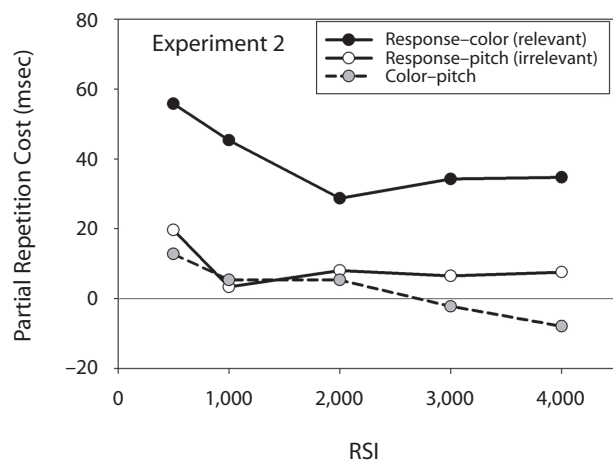
$p < .001$ ], indicating higher costs associated with color–response integration (41 msec) than with pitch–response integration (9 msec) and color–pitch integration (3 msec), as can be seen in Figure 3.

We also compared findings across the two experiments by running a mixed factors ANOVA on the partial repetition costs from Experiments 1 and 2 with RSI, effect type (relevant  $\times$  irrelevant stimulus feature, relevant stimulus feature  $\times$  response), and experiment as factors. The factor of experiment produced a main effect [ $F(1,100) = 14.91, p < .001$ ], due to higher average costs in Experiment 1 (34 msec) than in Experiment 2 (17 msec), 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 2 and 3: The costs associated with the relevant stimulus feature and the response were comparable across the two experiments (38 vs. 40 msec), whereas the two types of costs involving the irrelevant stimulus feature (pitch, in both cases) differed rather drastically. In fact, costs due to stimulus–stimulus interactions were more than 10 times larger in Experiment 1 than in Experiment 2 (35 vs. 3 msec), and the costs associated with pitch–response interactions were still about three times larger (28 vs. 9 msec).

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 tasks 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 feature-integration mechanisms by examining the temporal dynamics of unimodal and multimodal bindings, as



**Figure 3.** Partial repetition costs, as a function of response–stimulus interval (RSI), in Experiment 2.

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) and 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 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 4 sec. 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 3,000 msec. At first sight, the effects indicative of the integration of stimulus features seem to show marked differences: The unimodal auditory bindings in Experiment 1 are much more pronounced than are the audiovisual bindings in Experiment 2. However, note that the corresponding RSI functions are extremely similar (as also confirmed by the absence of an RSI  $\times$  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  $\times$  RSI function.

Our second guiding question concerned the role of attention in both 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 were not only almost the same in terms of intercept but virtually identical in terms of slope (see Figure 2). So whatever causes the decrease of effect size over time affects task-relevant and task-irrelevant features in the same way.

Our third question concerned the impact of the context or attentional set on unimodal 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 Figures 2 and 3 shows that the answer is somewhat complicated by what seems to be a mere floor effect (in Experiment 2) but is 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 with that in Experiment 1, does not seem to be due to faster decay but, rather, to a stronger point of departure.

Taken together, 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 when the features they involve are relevant to the task at hand (and are thus attended), such as when 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 N2-posterior-contralateral component (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 made by Talsma and Woldorff (2005), who found larger audiovisual 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 (2007) 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 is commonly done 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, 2007; Hommel & Colzato, 2004) and hints of a selective impact of attentional manipulations on retrieval (Hommel, Memelink, Zmigrod, & Colzato, 2009), 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 with the impact in 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 of our main observations: All interactions involving a task-relevant feature as defined above were sizable and rather robust across most of the RSIs, 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 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 on the retrieval of bindings, is regulated through a mechanism that operates on the intercept but not on 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, 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 only 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 corresponds to the interval that still allows for perceiving an object as constant. In Kahneman et al.'s (1992) original study, the delay between two presentations of the critical stimuli was in the range of milliseconds, but later studies found that object files can persist as long as 8 sec (Noles, Scholl, & Mitroff, 2005). It may be interesting to note that these rather long lifetimes were observed under conditions in which the interval between two presentations was filled by other visual events—events that served to bridge 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 for the participant's own response. From studies on causal perception, it 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 4 sec 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 the actual or functional lifetime of an episodic memory trace. This trace may or may not be active after having been created: It may be deleted after some time or, alternatively, may be rewritten or overwritten after a new combination of features appears (which would refer to its actual lifetime) or at least no longer be discriminable from alternative traces (which would refer to its functional lifetime) (Alvarez & Thompson, 2009). Note that our analyses focused on two particular representations (of S1 and S2), in which the features in question were combined in a particular way, but more than 100 combinations are coded in a typical sequential-effects experiment. If each single combination were stored for a longer period of 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 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 4 sec. 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 function. Under realistic circumstances, a certain time during which 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, in press; 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 at around 3–4 sec. It may be just 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 previous episode's likelihood of being retrieved.

A fourth alternative is that memory plays a role in the longer RSIs. Hommel and Colzato (2004) have provided some evidence that the integration of visual features takes several hundred milliseconds, and one can imagine that the RSI manipulation taps into an extended process comprising 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 information that is bound less accessible and/or somehow 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 of the relationship between memory processes and binding did not show hints of an interaction. For instance, Colzato, Raffone, and Hommel (2006) failed to find stronger binding effects for overlearned feature combinations; for instance, the binding between the pictures of a banana and the color yellow was no stronger than that between the pictures of a 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 codetermined by a number of factors, including the task relevance of the features involved; the 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.

## AUTHOR NOTE

Correspondence concerning this article should be addressed to S. Zmigrod, Cognitive Psychology Unit, Leiden University, Wassenaarseweg 52, 2333 AK, Leiden, The Netherlands (e-mail: szmigrod@fsw.leidenuniv.nl).

## REFERENCES

- AKYÜREK, E. G., TOFFANIN, P., & HOMMEL, B. (2008). Adaptive control of event integration. *Journal of Experimental Psychology: Human Perception & Performance*, *34*, 569-577. doi:10.1037/0096-1523.34.3.569
- ALLPORT, D. A., TIPPER, S. P., & CHMIEL, N. R. J. (1985). Perceptual integration and postcategorical filtering. In M. I. Posner & O. S. M. Marin (Eds.), *Attention & performance XI* (pp. 107-132). Hillsdale, NJ: Erlbaum.
- ALVAREZ, G. A., & THOMPSON, T. W. (2009). Overwriting and re-binding: Why feature-switch detecting tasks underestimate the binding capacity of visual working memory. *Visual Cognition*, *17*, 141-159. doi:10.1080/13506280802265496
- BERTELSON, P. (1963). S-R relationships and reaction times to new versus repeated signals in a serial task. *Journal of Experimental Psychology*, *65*, 478-484.
- BERTELSON, P., VROOMEN, J., DE GELDER, B., & DRIVER, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*, 321-332.
- BURKE, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology*, *4*, 121-138.
- COLZATO, L. S., RAFFONE, A., & HOMMEL, B. (2006). What do we learn from binding features? Evidence for multilevel feature integration. *Journal of Experimental Psychology: Human Perception & Performance*, *32*, 705-716. doi:10.1037/0096-1523.32.3.705
- DESCHÉPPER, B., & TREISMAN, A. [M.] (1996). Visual memory for novel shapes: Implicit coding without attention. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *22*, 27-47.
- DYSON, B. J., & QUINLAN, P. T. (2004). Stimulus processing constraints in audition. *Journal of Experimental Psychology: Human Perception & Performance*, *30*, 1117-1131. doi:10.1037/0096-1523.30.6.1117
- EINHORN, H. J., & HOGARTH, R. M. (1986). Judging probable cause. *Psychological Bulletin*, *99*, 3-19.
- GAO, T., & SCHOLL, B. J. (in press). Are objects required for object-files? Roles of segmentation and spatiotemporal continuity in computing object persistence. *Visual Cognition*. doi:10.1080/13506280802614966
- GOLDSTEIN, E. B. (2007). *Sensation and perception* (7th ed.). Belmont, CA: Thomson Wadsworth.
- GORDON, R. D., & IRWIN, D. E. (2000). The role of physical and conceptual properties in preserving object continuity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 136-150.
- GRUBER, H. E., FINK, C. D., & DAMM, V. (1957). Effects of experience on perception of causality. *Journal of Experimental Psychology*, *53*, 89-93.
- HALL, M. D., PASTORE, R. E., ACKER, B. E., & HUANG, W. (2000). Evidence for auditory feature integration with spatially distributed items. *Perception & Psychophysics*, *62*, 1243-1257.
- HOMMEL, B. (1996). The cognitive representation of action: Automatic integration of perceived action effects. *Psychological Research*, *59*, 176-186. doi:10.1007/BF00425832
- HOMMEL, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition*, *5*, 183-216. doi:10.1080/713756773
- HOMMEL, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*, 494-500. doi:10.1016/j.tics.2004.08.007
- HOMMEL, B. (2005). How much attention does an event file need? *Journal of Experimental Psychology: Human Perception & Performance*, *31*, 1067-1082. doi:10.1037/0096-1523.31.5.1067
- HOMMEL, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research*, *71*, 42-63. doi:10.1007/s00426-005-0035-1
- HOMMEL, B. (2009). Action control according to TEC (theory of event coding). *Psychological Research*, *73*, 512-526. doi:10.1007/s00426-009-0234-2
- HOMMEL, B. (2010). Grounding attention in action control: The intentional control of selection. In B. J. Bruya (Ed.), *Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 121-140). Cambridge, MA: MIT Press.
- HOMMEL, B., & COLZATO, L. S. (2004). Visual attention and the temporal dynamics of feature integration. *Visual Cognition*, *11*, 483-521. doi:10.1080/13506280344000400
- HOMMEL, B., & COLZATO, L. S. (2009). When an object is more than a binding of its features: Evidence for two mechanisms of visual feature integration. *Visual Cognition*, *17*, 120-140. doi:10.1080/13506280802349787
- HOMMEL, B., MEMELINK, J., ZMIGROD, S., & COLZATO, L. S. (2009). How information of relevant dimension controls the creation and retrieval of feature-response binding. Manuscript submitted for publication.
- HÖTTING, K., & RÖDER, B. (2004). Hearing cheats touch, but less in congenitally blind than in sighted individuals. *Psychological Science*, *15*, 60-64.
- HYUN, J.-S., WOODMAN, G. F., & LUCK, S. J. (2009). The role of attention in the binding of surface features to locations. *Visual Cognition*, *17*, 10-24. doi:10.1080/13506280802113894
- KAHNEMAN, D., TREISMAN, A. [M.], & GIBBS, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*, 175-219.
- LOGAN, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*, 492-527.
- LUCK, S. J., & HILLYARD, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 1000-1014.
- LUCK, S. J., & HILLYARD, S. A. (1995). The role of attention in feature detection and conjunction discrimination: An electrophysiological analysis. *International Journal of Neuroscience*, *80*, 281-297.
- MASSARO, D. W. (1987). Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum.
- MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746-748.
- MITROFF, S. R., & ALVAREZ, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin & Review*, *14*, 1199-1204.
- MITROFF, S. R., ARITA, J. T., & FLECK, M. S. (2009). Staying in bounds: Contextual constraints on object-file coherence. *Visual Cognition*, *17*, 195-211. doi:10.1080/13506280802103457
- MITROFF, S. R., SCHOLL, B. J., & WYNN, K. (2004). Divide and conquer: How object files adapt when a persisting object splits into two. *Psychological Science*, *15*, 420-425.
- MONDOR, T. A., HURLBURT, J., & THORNE, L. (2003). Categorizing sounds by pitch: Effects of stimulus similarity and response repetition. *Perception & Psychophysics*, *65*, 107-114.
- MORDKOFF, J. T., & HALTERMAN, R. (2008). Feature integration without visual attention: Evidence from the correlated flankers task. *Psychonomic Bulletin & Review*, *15*, 385-389. doi:10.3758/PBR.15.2.385
- NOLES, N. S., SCHOLL, B. J., & MITROFF, S. R. (2005). The persistence of object file representations. *Perception & Psychophysics*, *67*, 324-334.
- PRATT, J., & HOMMEL, B. (2003). Symbolic control of visual attention: The role of working memory and attentional control settings. *Journal of Experimental Psychology: Human Perception & Performance*, *29*, 835-845. doi:10.1037/0096-1523.29.5.835
- REED, P. (1992). Effect of a signalled delay between an action and outcome on human judgement of causality. *Quarterly Journal of Experimental Psychology*, *44B*, 81-100.
- REED, P. (1999). Role of a stimulus filling an action-outcome delay in human judgments of causal effectiveness. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 92-102.
- SAIKI, J. (2009). Functional roles of memory for feature-location binding in event perception: Investigation with spatiotemporal visual search. *Visual Cognition*, *17*, 212-231. doi:10.1080/13506280802280230
- SHAMS, L., KAMITANI, Y., & SHIMOJO, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*, 788.
- TAKEGATA, R., BRATTICO, E., TERVANIEMI, M., VARYAGINA, O., NÄÄTÄNEN, R., & WINKLER, I. (2005). Preattentive representation

- of feature conjunctions for concurrent spatially distributed auditory objects. *Cognitive Brain Research*, **25**, 169-179. doi:10.1016/j.cogbrainres.2005.05.006
- TALSMA, D., & WOLDORFF, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, **17**, 1098-1114.
- TREISMAN, A. M., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12**, 97-136.
- TREISMAN, A. [M.], & SATO, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 459-478.
- VROOMEN, J., BERTELSON, P., & DE GELDER, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, **63**, 651-659.
- VROOMEN, J., & DE GELDER, B. (2004). Temporal ventriloquism: Sound modulates the flash-lag effect. *Journal of Experimental Psychology: Human Perception & Performance*, **30**, 513-518.
- WESSINGER, C. M., VANMETER, J., TIAN, B., VAN LARE, J., PEKAR, J., & RAUSCHECKER, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of Cognitive Neuroscience*, **13**, 1-7.
- YI, D.-J., TURK-BROWNE, N. B., FLOMBAUM, J. I., KIM, M.-S., SCHOLL, B. J., & CHUN, M. M. (2008). Spatiotemporal object continuity in human ventral visual cortex. *Proceedings of the National Academy of Sciences*, **105**, 8840-8845. doi:10.1073/pnas.0802525105
- ZEKI, S., & BARTELS, A. (1999). Toward a theory of visual consciousness. *Consciousness & Cognition*, **8**, 225-259.
- ZMIGROD, S., & HOMMEL, B. (2009). Auditory event files: Integrating auditory perception and action planning. *Attention, Perception, & Psychophysics*, **71**, 352-362. doi:10.3758/APP.71.2.352
- ZMIGROD, S., SPAPÉ, M., & HOMMEL, B. (2009). Intermodal event files: Integrating features across vision, audition, tactition, and action. *Psychological Research*, **73**, 674-684. doi:10.1007/s00426-008-0163-5

#### NOTES

1. 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). Reasons for doing so (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. For instance, the partial repetition costs for the loudness  $\times$  pitch interaction at a given RSI would be  $PR_{loudness \times pitch} = (RT_{loudness\ repeated/pitch\ alternated} + RT_{loudness\ alternated/pitch\ repeated})/2 - (RT_{loudness\ repeated/pitch\ repeated} + RT_{loudness\ alternated/pitch\ alternated})/2$ . Partial repetition costs thus correspond to the two-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 means that the repetition effects of the two given features did not interact. A value greater than zero indicates a binding-type interaction of the sort described in the text.

(Manuscript received April 24, 2009;  
revision accepted for publication August 12, 2009.)