ORIGINAL ARTICLE

Intermodal event files: integrating features across vision, audition, taction, and action

Sharon Zmigrod · Michiel Spapé · Bernhard Hommel

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, Brattico, Tervaniemi, Varyagina, Näätänen, & Winkler 2005). However, real objects are rarely defined and 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 after effects. 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

S. Zmigrod (✉) · M. Spapé · B. Hommel
Department of Psychology, Cognitive Psychology Unit,
Leiden University Institute for Psychological Research
and Leiden Institute for Brain and Cognition, Postbus 9555,
2300 RB Leiden, The Netherlands
e-mail: szmigrod@fsw.leidenuniv.nl

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, 2008).

The result patterns observed in these studies rule 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 Fig. 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
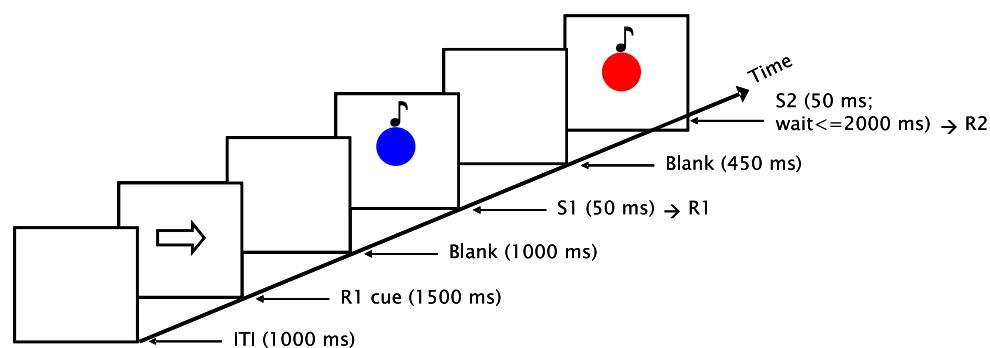
Fig. 1 Sequence of events in Experiment 1. A visual response cue signaled a left or right mouse button click (R1) that was to be delayed until presentation of an audiovisual stimulus S1 (S1 is used as a detection signal for R1). The audiovisual stimulus S2 appeared 450 ms after R1. S2 signaled R2, a speeded left or right mouse button click according to the instructed mapping and task

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, 2008). 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, 2008), 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 as 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.

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 Fig. 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, 2008), 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 (2008) 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 in. 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° 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 of two pure tones of 1,000 and 3,000 Hz 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 Fig. 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 (3,000 Hz) and the low pitch (1,000 Hz) 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 Fig. 1. A response cue with a right or left arrow appeared for 1,000 ms to signal R1, which was to be carried out as soon as S1 appeared. The duration between the response cue and S1 was 1,000 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 versus alternation of the response, and repetition versus alternation of the visual feature (color), and repetition versus alternation of the auditory feature (pitch).
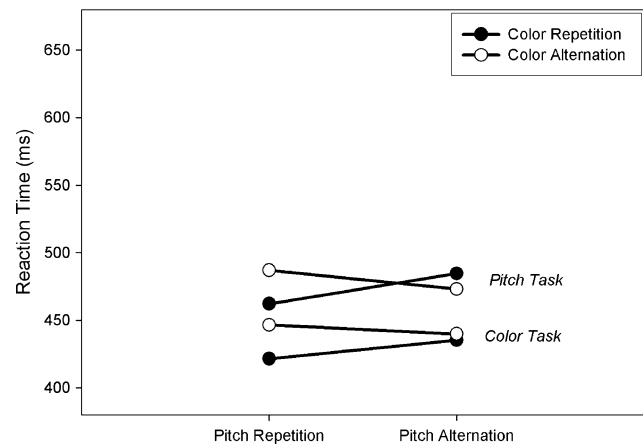
Results

Trials with incorrect R1 responses (1%), as well as missing (RT > 1,200 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

Table 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)

Attended modality	The relationship between the stimuli (S1 and S2)	Response			
		Repeated		Alternated	
		RT	PE	RT	PE
Visual	Color and pitch alternated	479	18.6	401	1.5
	Only color repeated	425	6.6	446	11.5
	Only pitch repeated	463	11.1	430	5.4
	Color and pitch repeated	399	2.8	443	14.5
Auditory	Color and pitch alternated	518	18.1	428	3.3
	Only color repeated	526	15.8	444	3.0
	Only pitch repeated	457	6.4	516	12.0
	Color and pitch repeated	430	3.1	494	19.6

Fig. 2 Reaction times of R2 in Experiment 1 as a function of repetition versus alternation of the stimuli (S1–S2) of visual feature color and auditory feature pitch, regardless of the response

R2 (see Table 1) were analyzed by means of four-way ANOVAs for repeated measures (see Table 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.

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 Fig. 2). Separate ANOVAs, split by task, revealed that it was more pronounced in, and statistically restricted to the pitch task (pitch task: $F(1,12) = 5.679$, $P < 0.05$; color task: $F(1,9) = 2.796$, ns),

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 Fig. 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

Table 2 Experiment 1: results of analysis of variance on mean reaction time (RT) of correct responses and percentage of errors (PE) of R2. $df = (1,12)$ for all effects

Effect	RT		PE	
	MSE	F	MSE	F
Task	87020.48	2.84	67.42	0.56
Response	7421.19	2.15	111.80	0.79
Pitch	776.48	0.46	9.31	0.16
Color	6000.87	3.53	0.17	0.01
Task × response	8.10	0.01	0.55	0.02
Task × pitch	6.39	0.00	22.58	0.43
Response × pitch	107254.79	71.26***	3739.88	35.17***
Task × response × pitch	42242.13	13.60**	819.81	13.48**
Task × color	907.23	0.33	6.64	0.38
Response × color	29501.07	25.51***	2228.02	10.99**
Task × response × color	21564.50	20.60***	573.84	6.84*
Pitch × color	10522.23	8.89**	76.47	1.04
Task × pitch × color	837.69	0.64	13.64	0.22
Response × pitch × color	532.61	0.15	14.51	0.35
Task × response × pitch × color	261.86	0.37	152.21	2.27

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$
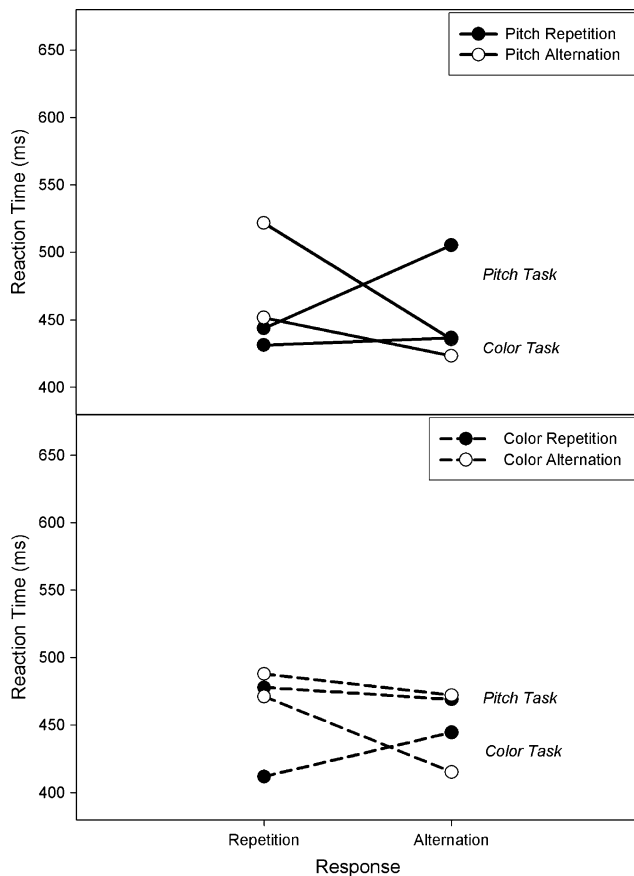
Fig. 3 Reaction times of R2 in Experiment 1 for repetition versus alternation of the stimuli in the auditory feature pitch and the visual feature color, as a function of response repetition (vs. alternation) and task

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: $F_{(1,12)} = 43.11$, $P < 0.0001$; auditory task: $F_{(1,12)} = 45.97$, $P < 0.0001$) and errors [visual task: $F_{(1,12)} = 12.55$, $P < 0.005$; auditory task: $F_{(1,12)} = 32.24$, $P < 0.0001$]. 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, $F_{(1,12)} = 4.89$, $P < 0.05$, and error rates, $F_{(1,12)} = 12.55$, $P < 0.005$; while no effects were obtained in the auditory task $F < 1$.

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 et al., 2003; Zmigrod & Hommel, 2008), 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 (2008), 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, 2008), 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, 2008), 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 possibility: 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 (1,000 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 500 ms, and the auditory features were based on 1,000 Hz 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 > 1,200 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 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 4.

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).
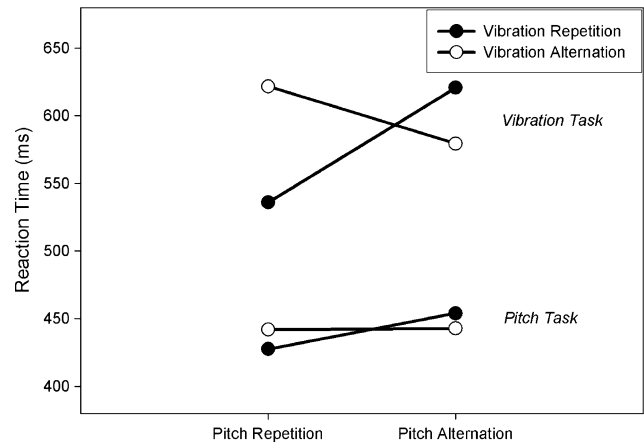
Stimulus integration. A significant interaction between pitch (repetition vs. alternation) and vibration rate (repetition vs. alternation) was obtained. This reflects a crossover pattern with slower responses for trials in which one feature repeats while the other alternates, as compared to complete

Table 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)

Attended modality	The relationship between the stimuli (S1 and S2)	Response			
		Repeated		Alternated	
		RT	PE	RT	PE
Auditory	Pitch and vibration alternated	478	7.8	407	5.2
	Only pitch repeated	483	6.6	425	1.9
	Only vibration repeated	407	2.4	477	8.2
	Pitch and vibration repeated	407	4.0	447	9.1
Tactile	Pitch and vibration alternated	608	19.8	551	5.8
	Only pitch repeated	611	15.7	630	11.0
	Only vibration repeated	639	15.4	604	12.7
	Pitch and vibration repeated	503	9.8	568	11.2

Fig. 4 Reaction times of R2 in Experiment 2 as a function of repetition versus alternation of the stimuli (S1–S2) of tactile feature vibration and auditory feature pitch, and task

repetitions or alternations (see Fig. 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).

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 Fig. 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, $F(1,9) = 17.14$, $P < 0.005$; response

by vibration in the vibration task, $F(1,9) = 26.51$, $P < 0.001$) but not for the task-irrelevant feature. In error rates, only the interaction between pitch and response repetition was reliable.

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

Table 4 Experiment 2: results of analysis of variance on mean reaction time (RT) of correct responses and percentage of errors (PE) of R2. $df = (1,9)$ for all effects

Effect	RT		PE	
	MSE	F	MSE	F
Task	875895.10	12.93**	1974.02	8.23*
Response	437.55	0.14	168.10	3.10
Pitch	12184.81	8.14*	0.62	0.02
Vibration	5699.80	3.32	40.00	0.84
Task × response	117.63	0.05	348.10	1.79
Task × pitch	607.04	0.37	18.22	0.62
Response × pitch	59354.31	12.41**	792.10	0.02*
Task × response × pitch	18432.21	7.33*	0.40	0.00
Task × vibration	4232.38	1.33	10.00	0.18
Response × vibration	15759.51	5.79*	70.22	0.56
Task × response × vibration	23149.33	10.29*	164.02	4.45
Pitch × vibration	58549.66	32.38***	0.90	0.02
Task × pitch × vibration	25819.86	11.03**	144.40	2.53
Response × pitch × vibration	219.70	0.16	9.02	0.40
Task × response × pitch × vibration	2822.15	0.82	27.22	0.32

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$
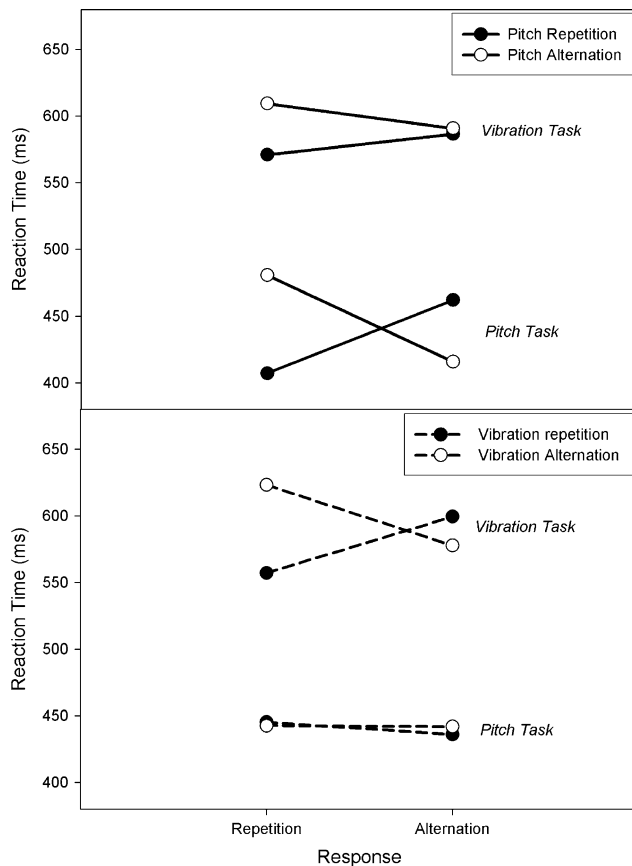
Fig. 5 Reaction times of R2 in Experiment 2 for repetition versus alternation of the stimuli in the auditory feature pitch and the tactile feature vibration, as a function of response repetition (vs. alternation) and task

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 (2008), 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 of the 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 suggests 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 et al., 2008). 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.

References

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.

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.

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 and Performance, 32*, 705–716.

Dutzi, I. B., & Hommel, B. (2008). The microgenesis of action-effect binding *Psychological Research*.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception & Performance, 27*, 229–240.

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a 'dimension weighting' account. *Perception & Psychophysics, 58*, 88–101.

Goldstein, E. B. (2007) (Ed.). *Sensation and perception* (*7th ed.*). Belmont, CA: Thomson Wadsworth.

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. (1993). The relationship between stimulus processing and response selection in the Simon task: Evidence for a temporal overlap. *Psychological Research, 55*, 280–290.

Hommel, B. (1998). Event files: evidences for automatic integration of stimulus-response episodes. *Visual Cognition, 5*, 183–216.

Hommel, B. (2004). Event files: feature binding in and across perception and action. *Trends in Cognitive Sciences, 8*, 494–500.

Hommel, B. (2005). How much attention does an event file need? *Journal of Experimental Psychology: Human Perception & Performance, 31*, 1067–1082.

Hommel, B., & Colzato, L. S. (2004). Visual attention and the temporal dynamics of feature integration. *Visual Cognition, 11*, 483–521.

Hommel, B., & Colzato, L. S. (2008). When an object is more than a binding of its features: Evidence for two mechanisms of visual feature integration. *Visual Cognition.*

Hommel, B., Memelink, J., Zmigrod, S., & Colzato, L. S. (2008). How information of relevant dimension control the creation and retrieval of feature-response binding, under revision.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral & Brain Sciences, 24*, 849–937.

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

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object specific information. *Cognitive Psychology, 24*, 175–219.

Leboe, J. P., Mondor, T. A., & Leboe, L. C. (2006). Feature mismatch effects in auditory negative priming: Interference as dependent on salient aspects of prior episodes. *Perception & Psychophysics, 68*, 897–910.

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

Mondor, T. A., Hurlburt, J., & Thorne, L. (2003). Categorizing sounds by pitch: Effects of stimulus similarity and response repetition. *Perception & Psychophysics, 65*, 107–114.

Posner, M. I., Nissen, J. J., & Klein, R. M. (1976). Visual dominance: An information processing account of its origins and significance. *Psychological Review, 83*, 157–171.

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

Sherrick, C. (1985). A scale for rate of tactual vibration. *Journal of the Acoustical Society of America, 78*, 78–83.

Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed audition objects. *Cognitive Brain Research, 25*, 169–179.

Treisman, A. M. (1996). The binding problem. *Current Opinion in Neurobiology, 6*, 171–178.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Van Erp, J. B. F., & Spapé, M. M. (2003). *Distilling the underlying dimensions of tactile melodies* (pp. 111–120). Dublin, Ireland: Eurohaptics 2003 proceedings.

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.

Wallace, M. T., Carriere, B. N., Perrault, T. J., Jr, Vaughan, J. W., & Stein, B. E. (2006). The development of cortical multisensory integration. *Journal of Neuroscience., 26*, 11844–11849.

Zmigrod, S., & Hommel, B. (2008). Auditory Event Files: Integrating auditory perception and action planning. *Perception & Psychophysics.*