Auditory event files: Integrating auditory perception and action planning

Sharon Zmigrod and Bernhard Hommel
Leiden University, Leiden, The Netherlands

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

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

Hommel (1998, 2004, 2005) extended Treisman's object file concept to include not only stimulus features, but also response-related feature information. A number of studies have provided evidence for this extension. In these studies, participants carried out two responses in a row. First, they were cued by a response cue signaling the first response, which, however, was carried out only after a visual trigger stimulus was presented. After 1 sec, another visual stimulus appeared, and the participants had to perform a binary-choice response to one of its features. As was expected, main effects of stimulus feature repetition were obtained. But more interestingly, stimulus and response repetition effects interacted: Repeating a stimulus feature sped up reaction time (RT) only if the response also repeated, whereas stimulus feature repetition slowed down RT if the response alternated. Apparently, stimulus

features were bound to response features, so that repeating one retrieved the other. This created conflict in partial repetition trials—that is, when the retrieved stimulus or response feature did not match the present one. Hence, facing a particular combination of stimulus and response features seems to create a multimodal *event file* (Hommel, 1998, 2004), which is retrieved if at least one of the features it includes is encountered again.

The existing theories in feature integration have been based largely on experiments using visual information, but it makes sense to assume that feature integration takes place in auditory perception as well. The auditory system allows us to perceive events based on the sound produced by them. And yet, an acoustic event is commonly made up of several features, among them pitch, timbre, loudness, and spatial position. Numerous studies have been done to look into how these features are perceived; however, in everyday life, we do not perceive features in isolation but, rather, perceive coherent, integrated acoustic events. Given that these features are processed in different areas of the auditory cortex (Kaas & Hackett, 1999; Wessinger et al., 2001), there should be a mechanism that integrates the auditory features into a coherent acoustic perception. Indeed, there is preliminary evidence for the existence of auditory binding. For instance, Hall et al. (2000) examined auditory feature integration of spatially distributed musical tones by having participants search for either a cued conjunction of pitch and timbre or a single cued value (pitch or timbre) in arrays of simultaneous tones in different lateralized positions. Their finding revealed more frequent illusory conjunctions when pitch and timbre features were separately presented, suggesting that, like the visual system, the auditory system differentiates the auditory features from the sound field and

S. Zmigrod, szmigrod@fsw.leidenuniv.nl

then integrates them according to their source. The investigators concluded that the auditory system binds its features with reference to their location, just as FIT (Treisman & Gelade, 1980) assumes to occur for the visual system. In addition, Leboe, Mondor, and Leboe (2006), who investigated different sources of auditory negative priming effects, found that repeated sounds in opposite locations were categorized more slowly than repeated sounds in the same location. In the interdomains of auditory perception and action, Mondor, Hurlburt, and Thorne (2003) found interactions between pitch and response repetition effects, which may indicate the integration of sound features and action.

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

Although there is ample evidence for the existence of event files in and across visual perception and action planning, the event file concept has not been systematically applied to auditory perception and action planning. Only a few studies have examined the binding mechanism in the auditory modality, and there is contradictory evidence regarding the role of attention in this mechanism. The aim of the present study was to investigate the feature binding mechanism in and across auditory perception and action planning. More specifically, we addressed three research issues: whether evidence for feature integration

in a standard object file can be observed for different auditory dimensions; whether evidence for stimulus–response integration effects can be obtained between the auditory modality and action planning; and whether these integration effects rely on or are mediated by attention.

EXPERIMENT 1

Experiment 1 was performed to determine whether auditory features are integrated into a coherent object representation and whether response-related features are also integrated with auditory features to produce an event file similar to Hommel's (1998, 2004) findings in the visual domain. The task followed Hommel's (1998) design, except that the stimuli were pure tone sounds. Participants were cued to prepare a response (left or right mouse button click), which they carried out (R_1) after the first stimulus (S_1). One second later, the second sound (S_2) was played, and the participants had to respond to the value of the relevant auditory feature by carrying out response R_2 (left or right mouse button click; see Figure 1).

The auditory features that were chosen for this experiment were pitch and loudness. Neuhoff, Kramer, and Wayand (2002) demonstrated that pitch and loudness have an interactive effect—that is, changes in one of these dimensions influence the other. On the basis of these results and the object file concept, we hypothesized that pitch and loudness features of S_1 are integrated and are still bound when S_2 is processed. If so, repeating the feature in one dimension should produce better performance if the feature in the other dimension is also repeated, whereas alternating the feature in one dimension should produce better performance if the feature in the other dimension is also alternated. In addition, we hypothesized that the features making up S_1 are integrated with R_1 and are still bound to it when S_2 is responded to, on the basis of the suggested event-file mechanism that posits that a specific combination of stimulus and response creates an episodic trace that is retrieved in case of any feature repetition (Hommel, 1998, 2004). If so, a response to S_2 should be better with a complete match or a complete mismatch between the previous response and a given auditory feature than with par-
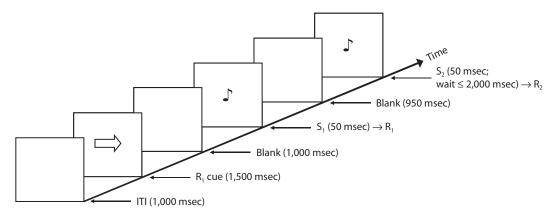

Figure 1. Sequence of events in the experiments. A response cue signaled a left or right mouse button click (R_1) that was to be delayed until presentation of the first stimulus (S_1). S_2 appeared 1,000 msec later. S_2 signaled R_2, a speeded left or right mouse button click, according to the task.

tial matches. Moreover, previous observation showed that pitch repetition interacts with response repetition (Mondor et al., 2003). To investigate the role of attention in auditory feature integration, we manipulated the feature that was relevant for responding to S_2. In one block of trials, only one of the two auditory features (pitch or loudness) was relevant, whereas in another block the other auditory feature was relevant. The task relevance of S_2 features (and the amount of attention consequently devoted to them) has been shown to affect the size of integration-related effects with visual stimuli (e.g., Hommel, 1998), and we were interested to see whether it would also modify such effects with auditory stimuli.

Method

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

Apparatus and Stimuli. The experiment was controlled by a Targa Pentium 3, attached to a Targa TM 1769-A 17-in. monitor. The participants faced the monitor at a distance of about 60 cm. The loudspeakers were located on both sides of the screen (approximately 25°) at a distance of 70 cm. The stimuli S_1 and S_2 were composed from two pure tones of 1000 and 3000 Hz with a duration of 50 msec and were presented at 65 and 75 dB SPL, respectively. Visual response cues were presented in the middle of the screen (see Figure 1), with a right or left arrow indicating a right or left response (R1), respectively. Responses were made by clicking on the left or the right mouse button with the index and middle fingers of the dominant hand.

Procedure and Design. The experiment was composed of two sessions: In one session, pitch was the relevant dimension for the task, and the participants had to respond to whether pitch was high or low; in the other session, loudness was the relevant dimension for the task, and the participants had to respond to whether loudness was high or low. The sessions were counterbalanced between participants. Each session contained a practice block with 10 practice trials and an experimental block with 128 experimental trials. The order of the trials was randomized. The participants had to carry out two responses per trial. R_1 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 S_1 appeared, regardless of its pitch or its loudness. R_2 was a binary-choice reaction to S_2. In the pitch-relevant session, half of the participants responded to the high pitch (3000 Hz) and the low pitch (1000 Hz) by pressing on the left and right mouse buttons, respectively, whereas the other half received the opposite mapping. In the loudness-relevant task, half of the participants responded to the loud sound (75 dB SPL) and to the soft sound (65 dB SPL) by pressing on the left and right mouse buttons, respectively, whereas the other half received the opposite mapping. The participants were asked to respond as quickly and accurately as possible.

The sequence of events in each trial is shown in Figure 1. A response cue with a right or left arrow visually presented for 1,500 msec signaled response R_1, which was to be carried out after S_1 had been played. S_2 was played 1 sec after the response to S_1, with the pitch (in the pitch session) or loudness (in the loudness session) signaling the second response (R_2). In case of incorrect or absent responses, an error message was presented. R_2 speed (RT) and accuracy (percentage of errors, or PE) were analyzed for all the trials with correct R_1 responses as a function of session (pitch/loudness), repetition versus alternation of the response, and repetition versus alternation of the stimulus dimensions pitch and loudness.

Results

Trials with incorrect R_1 responses (1.7%), as well as missing or anticipatory (RT < 100 msec) R_2 responses (0.7%), were excluded from analysis. The mean RT for R_1 was 270 msec ($SD = 88$). From the remaining data, mean RTs and PEs for R_2 were analyzed as a function of the four variables: the task-relevant stimulus feature (loudness vs. pitch), or *task* for short; the relationship between the responses R_1 and R_2 (alternation vs. repetition); the relationship between S_1 and S_2 on the pitch dimension (alternation vs. repetition); and the relationship between S_1 and S_2 on the loudness dimension (alternation vs. repetition). ANOVAs were performed by using a four-way design for repeated measures. Table 1 provides an overview of the RT and PE means obtained for R_2 performance.

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

Second, we will address the *stimulus integration effect* by examining the interactions between repetition and alternation of the stimulus features: There was an interaction between pitch repetition (vs. alternation) and loudness repetition [$F(1,11) = 11.07$, $p < .01$], indicating that, with pitch repetition, performance was quicker if loudness was also repeated than if loudness alternated, whereas with pitch alternation, performance was quicker if loudness alternated than if it was repeated (see Figure 2). This result provides support for auditory feature integration between pitch and loudness.

Third, we will consider *stimulus–response integration effects* by examining the interactions between repetition and alternation of the response and the stimulus features. There were interactions between response repetition and

Table 1
Experiment 1: Means and Standard Errors of Mean Reaction Times (RTs, in Milliseconds) and Percentages of Errors (PEs) for Responses to the Second Stimulus As a Function of the Attended Dimension, the Relationship Between the Stimuli (Repetition vs. Alternation), and the Relationship Between the Responses (Repetition vs. Alternation)

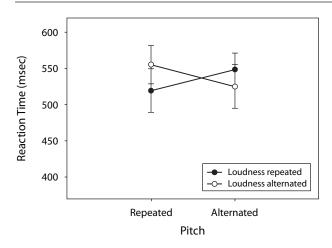| Attended Dimension | Stimulus Feature Repeated | Response | | | | | | | |
		Repeated				Alternated			
		RT		PE		RT		PE	
		M	SE	M	SE	M	SE	M	SE
Loudness	Neither	553	41	8.8	3.0	472	31	4.1	1.9
	Loudness	557	24	12.4	2.8	553	36	9.8	3.7
	Pitch	553	31	11.0	4.5	517	27	3.8	2.6
	Both	486	32	6.3	3.2	541	35	15.8	4.9
Pitch	Neither	574	41	11.3	2.2	502	39	5.6	2.3
	Loudness	564	38	14.6	3.9	521	40	8.2	2.7
	Pitch	548	42	19.1	5.3	604	38	18.4	5.5
	Both	507	42	7.9	2.7	545	45	21.6	4.5

Figure 2. Reaction times in Experiment 1 as a function of repetition versus alternation of pitch and loudness.

pitch repetition in RTs [$F(1,11) = 42.45$, $p < .0001$] and PEs [$F(1,11) = 8.90$, $p < .05$], showing that response repetition facilitates performance if the pitch repeats but impairs performance if the pitch alternates. Furthermore, there was an interaction between response repetition and loudness repetition in RTs [$F(1,11) = 5.14$, $p < .05$] and PEs [$F(1,11) = 9.30, p < .05$], showing that the responses were faster and more accurate for total repetition or total alternation of the response and the loudness than for partial repetition. In addition, a three-way interaction among task, response, and loudness in RTs [$F(1,11) = 6.63$, $p < .05$] was obtained, indicating sensitivity to the task-relevant feature in this stimulus–response effect. Separate ANOVAs confirmed that response interacted significantly in RTs only with loudness in the loudness task [$F(1,11) = 7.38$, $p < .05$] and not in the pitch task ($F < 1$). These interactions show stimulus–response effects between the response and the auditory stimuli. In the case of loudness, it was modulated by task relevance (see Figure 3).

Discussion

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

Our findings seem consistent with those of a recent auditory study by Mondor and Leboe (2008). These authors observed that the impact of pitch repetition on tone detection performance depends on response repetition, which seems to fit with our present stimulus–response integration effects. In particular, they found pitch repetition benefits if both the prime and the probe tone were to be detected and, thus, accompanied by the same response, and pitch-

repetition costs if the prime was to be ignored and, thus, not accompanied by a response. This outcome pattern bears similarities with our findings: good performance if both pitch and response repeat or both alternate, but bad performance if one repeats but not the other. However, Mondor and Leboe manipulated the response requirements between participants, which may have induced different attentional sets and strategies in the two tasks. For instance, ignoring primes in a detection task may lead to *inhibition of return* (Posner & Cohen, 1984), which may explain stimulus repetition costs without referring to response requirements; indeed, ignoring the prime and omitting a response to it led to a 70-msec *increase* in RT. Accordingly, it is not clear whether the observations of Mondor and Leboe reflect the same mechanisms that underlie the pitch × response interactions obtained in the present study.

Our findings reveal an interesting dissociation between the integration of stimulus features and the integration of stimulus and response features—a dissociation in which attention induced by task relevance plays a major role. Stimulus–response integration seems to be restricted mainly to the stimulus features that are task relevant: pitch in the pitch task and loudness in the loudness task. In contrast, different features of the same stimulus seem to be integrated irrespective of task-relevance, as evidenced by the reliable interaction between pitch and loudness under conditions that rendered only one of them relevant at any given time.
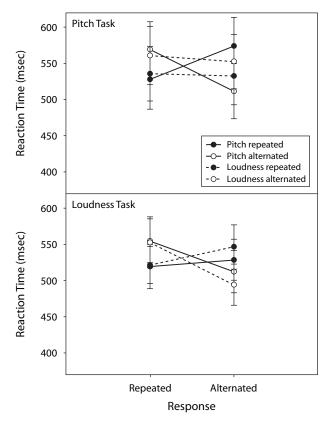

Figure 3. Reaction times in Experiment 1 for the repetition versus alternation of relevant and irrelevant stimuli (pitch and loudness) as a function of response (repetition vs. alternation) in pitch and loudness tasks.

One possible explanation might be that the physical attributes of the features influence one another; that is, loudness is known to be affected by frequency and pitch by intensity. It is also interesting to see that, in stimulus–response integration, task relevance was more effective in excluding irrelevant loudness information than irrelevant pitch information. In other words, in the present study, loudness was more sensitive to task relevance than was pitch.

We think that all these aspects of our findings point to the same integration principle: Features of events (whether they refer to stimuli or responses) are integrated to the degree that the activations of their codes overlap in time. This principle underlies the concept of conditioning (Pavlov, 1927) and seems crucial for the hippocampal integration of episodic stimulus and action events (Bangasser, Waxler, Santollo, & Shors, 2006). First, consider the respective roles that this principle plays in the integration of stimulus features versus the integration of stimulus and response features. As is indicated in Figure 4A, the activations of stimulus feature codes are likely to overlap in time even if they are peaking at different time points—that is, even if stimulus features are registered asynchronously. Accordingly, they are likely to be bound to each other, thus producing a partial-overlap cost. However, the earlier a feature is coded, the earlier its code decays, suggesting that quickly coded features are less likely to overlap in time with response code activation. In our study, we found that RTs were shorter in the loudness task than in the pitch task (see Figure 3), probably due to the greater saliency of loudness and/or the better discriminability of the loudness values we chose, suggesting that, in this experiment, loudness was coded more quickly than pitch.[1] With respect to the temporal relations depicted in Figure 4, this implies that response code activation started earlier, in our experiment, in the loudness task than it did in the pitch task. On top of that, there is evidence that loudness codes decay more quickly than pitch codes (Clement, Demany, & Semal, 1999), which would further work against the integration of loudness and response. We can thus conclude that the code overlap principle accounts for both the observation that task relevance did not affect stimulus integration and the finding that it did affect stimulus–response integration.

Making a feature dimension relevant to a task is likely to increase the weights (or gain) of that dimension's codes (Bundesen, 1990; Found & Müller, 1996; Hommel, Müsseler, Aschersleben, & Prinz, 2001), which again may result in stronger and/or more enduring activation (see Figure 4B). This means that task-relevant features induce activations that are more likely to overlap with the response activation. As a consequence, task-relevant features should be more likely to be integrated with the response than should task-irrelevant features, just as we observed in Experiment 1.

EXPERIMENT 2

Experiment 1 suggests that pitch and loudness are spontaneously integrated both with each other and with the response, at least if the given feature is task relevant. In Experiment 2, we investigated whether these observations can be extended to stimulus location. Many authors have emphasized the possibly crucial role of stimulus location in feature integration (in vision, Treisman & Gelade, 1980; in audition, Hall et al., 2000; Leboe et al., 2006).
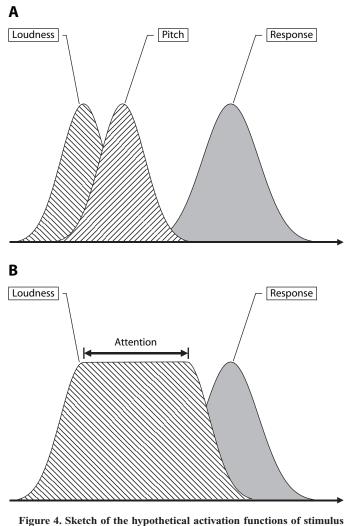
On the one hand, this could mean that spatial location is so important for feature integration that it does not matter whether location information is nominally relevant or irrelevant for a given task. This would still be consistent with the feature overlap principle, assuming that location features are strongly weighted irrespective of the task, but it would imply that the proposed relationship between task relevance and weighting does not apply to location. On the other hand, however, it is true that many tasks that are taken to demonstrate the crucial role of location have used spatial responses. Assuming that responses are represented, prepared, and planned in terms of their perceptual features (Hommel, 1996; Hommel et al., 2001), it is possible that defining a response set in terms of spatial features (e.g., by characterizing responses as *left* and *right*) attracts attention to the spatial dimension(s) and, thus, induces a stronger weighting of spatial codes. Indeed, Fagioli, Hommel, and Schubotz (2007) found evidence that preparing for particular types of actions (grasping vs. pointing) attracts attention to the features that are relevant for defining these actions (size vs. location). Along the same lines, Hommel (2007) observed that the integration of visual stimulus location and the response is much more pronounced when the response alternatives are spatially defined (left vs. right) than when they are not (pressing a key once vs. twice). Hence, it is possible that the previous findings of integration of (nominally) irrelevant location information and the response reflect not so much a central role of stimulus location in feature integration as the fact that defining responses spatially makes location task-relevant.

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

Method

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

Procedure and Design. The procedure was the same as that in Experiment 1, with the following exceptions. The loudspeakers were placed at an upper and lower position at 45º from the center of the screen. The stimuli S_1 and S_2 were composed from two pure tones of

A

B

Figure 4. Sketch of the hypothetical activation functions of stimulus codes. (A) In our experiment, loudness was coded more quickly than pitch was, so that the activation of pitch codes (even as the irrelevant dimension) was more likely to overlap with response code activation. (B) Task relevance of a given feature increased the duration of code activation, so that even codes that were activated early in time now overlapped with response code activation.

1000 and 3000 Hz, with durations of 50 msec, presented at approximately 70 dB SPL. The experiment was composed of two sessions: In one session, pitch was relevant for responding to S_2; in the other session, location was relevant to S_2, requiring a response to the top versus bottom location. The sessions were counterbalanced between participants. Each task contained a practice block with 15 practice trials and an experimental block with 96 experimental trials. The order of the trials was randomized.

The spatial response set group saw a left or right arrow indicating a left or right mouse click, respectively; responses to S_1 and to S_2 were made as in Experiment 1. The nonspatial response set group saw the word GO or NO GO, indicating whether to emit or withhold the response, respectively. Responses on the GO trials were made by clicking on the left mouse button; the NO-GO trials for S_1 lasted 500 msec.

Results and Discussion

Trials with incorrect R_1 responses (1%) as well as missing or anticipatory R_2 responses (RT < 100 msec: 0.1%)

were excluded from the analysis. The mean RTs for R_1 were 330 msec ($SD = 78$) for the spatial response set group and 341 msec ($SD = 114$) for the nonspatial response set group. From the remaining data, mean RTs and PEs for R2 were analyzed as a function of five variables: the task (pitch vs. location as the relevant S_2 feature), the relationship (repetition vs. alternation) between S_1 and S_2 with regard to pitch and location, the relationship (repetition vs. alternation) between responses R_1 and R_2, and the response set (spatial vs. nonspatial) (see Table 2 for mean RTs and PEs). ANOVAs were performed by using a mixed design with repeated measures on four variables and with response set as a between-groups variable (see Table 3 for the outcomes).

Let us consider the outcomes according to their theoretical implications. First, we will address the *task effects* that reflect the impact of the task on the stimulus dimen-

Table 2
Experiment 2: Means and Standard Errors of Mean Reaction Times (RTs, in Milliseconds) and Percentages of Errors (PEs) for Responses to the Second Stimulus As a Function of the Response Set (Spatial or Nonspatial), the Attended Dimension, the Relationship Between the Stimuli (Repetition vs. Alternation), and the Relationship Between the Responses (Repetition vs. Alternation)

Response Set	Attended Dimension	Stimulus Feature Repeated	Response							
			Repeated				Alternated			
			RT		PE		RT		PE	
			M	SE	M	SE	M	SE	M	SE
Spatial	Location	Neither	496	33	12.8	3.1	453	23	1.8	1.4
		Location	534	28	17.1	2.8	542	24	7.6	2.7
		Pitch	506	30	8.0	2.3	496	25	7.9	2.4
		Both	443	22	5.4	2.0	505	26	11.3	2.7
	Pitch	Neither	502	30	15.0	2.2	437	29	4.9	2.9
		Location	474	29	12.0	3.2	470	30	13.6	2.9
		Pitch	508	30	11.8	3.1	482	30	9.0	2.6
		Both	426	26	8.2	2.3	513	29	13.1	4.2
Nonspatial	Location	Neither	432	32	15.0	3.0	383	22	7.4	1.3
		Location	480	28	11.9	2.7	453	23	13.3	2.6
		Pitch	448	29	11.8	2.2	420	24	9.0	2.3
		Both	396	21	8.2	1.9	410	25	13.1	2.6
	Pitch	Neither	417	29	11.7	2.2	396	28	11.0	2.8
		Location	452	28	10.8	3.1	388	29	8.9	2.8
		Pitch	436	29	7.1	3.0	480	29	14.0	2.5
		Both	387	25	7.2	2.2	406	28	14.4	4.0

sions and the response. Second, we will consider the *stimulus integration effects*; these effects are revealed by interactions between the stimulus features, showing that repetition of a particular feature enhances performance if the other feature is also repeated and hinders performance if the other feature is alternated. Third, we will discuss *stimulus–response integration effects* by examining the interactions between repetition and alternation of the response and the stimulus features. Finally, we will address *response set effects*.

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

Stimulus integration effects. Pitch repetition interacted with location repetition, reflecting the standard crossover pattern with slower responses for trials in which one feature repeats while the other alternates; interestingly, it was more prominent when the relevant feature was repeated rather than alternated, which may point to the role of attention in the process (see Figure 5). This interaction was also modified by task, suggesting that the pitch × location interaction was somewhat more pronounced in the location task than in the pitch task, but it was clearly reliable in both [$F(1,27) = 66.44$, $p < .0001$, and $F(1,27) = 16.54$, $p < .0001$, respectively].

The latter observation seems inconsistent with the findings of Mondor and Leboe (2008), who failed to obtain interactions between pitch and location repetition when using a nonspatial response set. However, as was pointed out earlier, they used a detection task that did not require the discrimination of any stimulus feature. This design choice was likely to prevent feature bindings from affecting performance in several ways. For one, it yielded average RTs of less than 300 msec, which may have been too short to allow for the complete retrieval of the binding from the previous trial. Indeed, when Mondor and Leboe shortened the interval between the prime and the probe— a manipulation that they considered would facilitate binding retrieval and that effectively increased RTs—a close-to-significant interaction between pitch and location repetition effects was obtained. Moreover, a detection task is likely to induce rather shallow perceptual-coding processes, which again is likely to hamper the feature-matching process necessary to retrieve a particular binding. In any case, the present findings suggest that evidence for pitch–location binding can be obtained under favorable conditions.

To summarize, we were able to extend our observation of spontaneous pitch–loudness integration from Experiment 1 to the integration of pitch and location. Again, features from the two auditory dimensions involved were bound even though only one dimension was relevant at a time, suggesting that the mere temporal overlap of code activation is sufficient for integration.

Stimulus–response integration effects. Analogously to Experiment 1, pitch and location repetition entered two-way interactions with response repetition, in both RTs and PEs, reflecting worse performance if a stimulus-feature repetition was accompanied by an alternation of the re-

Table 3
Results of ANOVA on Mean Reaction Times (RTs) of
Correct Responses and Percentages of Errors (PEs) for
the Second Response in Experiment 2

Effect	RT			PE		
	MS_e	F	p	MS_e	F	p
Response set (S)	456,360.06	3.89	.059	373.38	0.79	.381
Task (T)	23,002.41	0.96	.335	1.99	0.01	.909
Response (R)	5,032.01	1.36	.254	0.03	0.00	.989
Pitch (P)	1,022.49	0.49	.491	95.27	2.40	.133
Location (L)	56.20	0.02	.887	297.88	3.97	.056
T × R	816.12	0.17	.683	647.12	7.67**	.010
T × P	28,779.30	13.54***	.001	46.96	0.58	.453
T × L	33,450.01	11.16**	.002	3.25	0.06	.816
R × P	82,731.54	33.22***	.000	1,730.53	17.32***	.000
T × R × P	7,570.75	3.25	.082	26.24	0.34	.566
R × L	39,527.31	15.10***	.001	1,153.42	13.44***	.001
T × R × L	2,687.12	1.39	.249	2.47	0.05	.824
P × L	145,149.32	73.60***	.000	69.45	1.30	.264
T × P × L	10,938.12	5.18*	.031	56.30	1.01	.323
R × P × L	5,761.72	3.35	.078	108.75	1.25	.273
T × R × P × L	403.56	0.21	.649	37.01	0.49	.489
T × S	4,988.43	0.21	.651	23.15	0.15	.698
R × S	6,499.33	1.76	.196	94.31	0.66	.423
P × S	47.70	0.02	.881	7.82	0.20	.660
L × S	2,038.26	0.75	.394	255.25	3.40	.076
T × R × S	4,101.46	0.86	.363	18.92	0.22	.640
T × P × S	28.65	0.01	.908	1.30	0.02	.900
R × P × S	40.02	0.02	.900	59.32	0.59	.448
T × R × P × S	836.78	0.36	.554	271.64	3.49	.073
T × L × S	550.30	0.18	.672	34.54	0.58	.451
R × L × S	40,724.50	15.55***	.001	164.97	1.92	.177
T × R × L × S	14,871.09	7.70**	.010	641.79	13.10***	.001
P × L × S	3,782.71	1.92	.177	93.51	1.75	.197
T × P × L × S	4,032.12	1.91	.178	0.29	0.01	.943
R × P × L × S	521.03	0.30	.586	103.76	1.20	.284
T × R × P × L × S	639.43	0.33	.568	4.68	0.06	.805

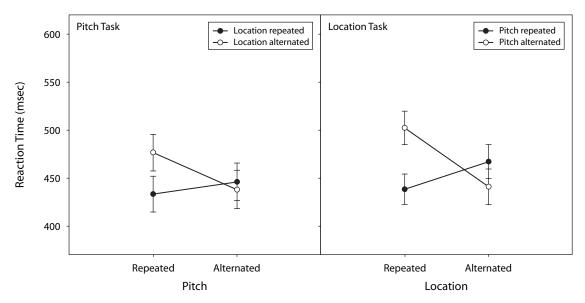Note—$df = (1,27)$.   *$p < .05$.   **$p < .01$.   ***$p < .001$.

Figure 5. Reaction times in Experiment 2 as a function of repetition versus alternation of pitch and location in the pitch task (left panel) and the location task (right panel).
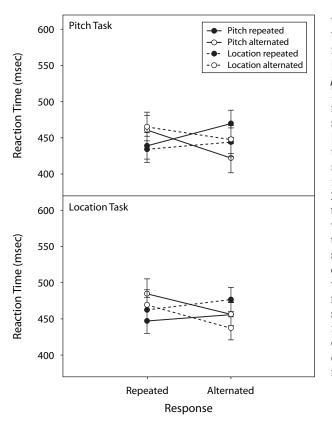
Figure 6. Reaction times in Experiment 2 in the pitch task (upper panel) and the location task (lower panel) for relevant and irrelevant stimuli (repetition vs. alternation) as a function of response (repetition vs. alternation).

sponse or vice versa (see Figure 6). The pitch × response interaction was unaffected by task, and a separate analysis confirmed that it was still reliable in the location task for both RTs [$F(1,27) = 9.04$, $p < .01$] and PEs [$F(1,27) = 12.70$, $p < .001$], as well as in the pitch task for both RTs [$F(1,27) = 27.10$, $p < .0001$] and PEs [$F(1,27) = 7.25$, $p < .05$]. The location × response interaction was also unaffected by task. A reliable effect between location and response was observed in the pitch task for both RTs [$F(1,27) = 4.3$, $p < .05$] and PEs [$F(1,27) = 8.614$, $p < .01$], as well as in the location task for both RTs [$F(1,27) = 15.41$, $p < .001$] and PEs [$F(1,27) = 8.56$, $p < .01$].

Response set effects. The response set manipulation did not yield a reliable main effect in RTs or PEs, even though participants tended to respond more quickly with a nonspatial than with a spatial set—presumably reflecting the reduced response uncertainty in the nonspatial go/no-go task. There were two reliable effects: The interaction between location and response was modified by response set (in RTs), and this three-way interaction was further modified by task (in RTs and PEs). Separate ANOVAs revealed that the location × response interaction was reliable only in the spatial response set condition [$F(1,13) = 39.43$, $p < .0001$] and not in the nonspatial response set condition ($F < 1$), indicating stronger activation when the response included spatial features. Moreover, to disentangle

the contributions to the four-way interaction, we analyzed the two tasks separately. In the location task, location and response repetition interacted significantly both in RTs [$F(1,27) = 15.41$, $p < .001$] and in PEs [$F(1,27) = 8.52$, $p < .01$], with no modulation by response set (see Figure 7, lower panel). However, in the pitch task, the location × response interaction was further modified by response set both in RTs [$F(1,27) = 20.86$, $p < .0001$] and in PEs [$F(1,27) = 9.94$, $p < .005$]. Separate analyses of the pitch task by response set revealed significant location × response interactions only for the spatial response set both in RTs [$F(1,13) = 19.16$, $p < .001$] and in PEs [$F(1,13) = 24.47$, $p < .0001$; see Figure 7, upper panel] and not in the nonspatial response set ($F < 1$). This pattern is in line with our expectation that a spatial response set amounts to making location task relevant, even with respect to stimulus coding. If location is task relevant by requiring discrimination of S_2 locations, location codes are strongly weighted anyway. As a consequence, stimulus location and responses are integrated, no matter whether the response set is spatially defined or not. However, when location is irrelevant with regard to S_2 (i.e., in the pitch task), location codes are weighted strongly only if location is relevant for discriminating the two responses, but not if a nonspatial response set is used.
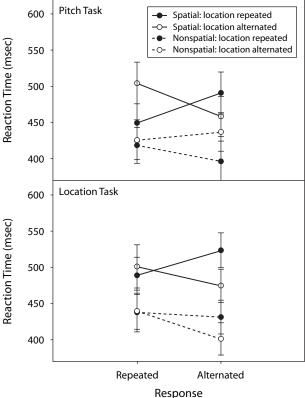

Figure 7. Reaction times in Experiment 2 in the pitch task (upper panel) and the location task (lower panel) for stimulus location (repetition vs. alternation) as a function of response (repetition vs. alternation) and response set group (spatial vs. nonspatial).

GENERAL DISCUSSION

The aim of our study was to investigate the binding mechanism in and across auditory perception and action. In both experiments, we found evidence for the spontaneous integration of auditory features: pitch and loudness in Experiment 1 and pitch and location in Experiment 2. Even though our participants were not instructed or required to create any feature conjunction, and even though nothing could be gained by doing so, the features of S_1—a mere go signal—were apparently integrated into a coherent representation. This outcome is in line with previous findings in visual perception, where feature integration effects were obtained between shape and color or shape and location (Hommel, 1998; Hommel & Colzato, 2004), and with findings from auditory studies, where evidence of integration was found for pitch and timbre (Hall et al., 2000; Takegata et al., 2005). We can conclude that feature-binding processes are not restricted to visual object perception, the modality targeted by FIT (Kahneman et al., 1992; Treisman & Gelade, 1980), but follow comparable principles in integrating auditory information.

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

The fact that evidence for feature integration processes was obtained even under conditions in which the processing of simple features would be sufficient supports the idea that integration occurs rather automatically. And yet, which information is integrated seems to be determined by the action goal. In particular, features that vary on dimensions that are relevant for defining a target stimulus or a response alternative are more likely to become part of bindings than are features unrelated to such dimensions. Another principle underlying feature integration seems to be a temporal overlap of code activation. Codes of stimulus features seem to be processed sufficiently close in time to produce overlapping activations, even if the time needed to process them differs (see Experiment 1) and even if only one of them is task relevant. That is, features belonging to the same physical stimulus are likely to become part of the same object file. The integration of stimuli and responses is more sensitive to temporal char-

acteristics. Stimuli that are closer in time to execution of a response seem to be more likely to be integrated with it. This fits with earlier observations of Hommel (2005), who found stimulus–response integration for stimuli presented briefly before, concurrently with, or even after the execution of the corresponding response, but no integration for stimuli presented during the planning of that response. Apparently, then, response execution provides the information necessary to trigger the integration process. A plausible candidate for pulling the trigger is the success of the response, which may signal that integrating the response with the apparently suitable context conditions is useful (Schultz, 2002). This possibility is strengthened by the finding that the integration of visual stimulus features and manual responses in a task such as ours is facilitated by presenting positively toned pictures after the execution of R_1 (Colzato, van Wouwe, & Hommel, 2007). Taken together, our findings provide evidence for the existence of temporary feature binding in auditory perception and action, suggesting a general principle of how events are cognitively represented—presumably, in terms of *event files*, as proposed by Hommel (1998, 2004).

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 and performance XI* (pp. 107-132). Hillsdale, NJ: Erlbaum.

BANGASSER. D. A., WAXLER, D. E., SANTOLLO, J., & SHORS, T. J. (2006). Trace conditioning and the hippocampus: The importance of contiguity. *Journal of Neuroscience*, **26**, 8702-8706. doi:10.1523/JNEUROSCI.1742-06.2006

BUNDESEN, C. (1990). A theory of visual attention. *Psychological Review*, **97**, 523-547.

CLEMENT, S., DEMANY, L., & SEMAL, C. (1999). Memory for pitch versus memory for loudness. *Journal of the Acoustical Society of America*, **106**, 2805-2811.

COLZATO, L. S., VAN WOUWE, N. C., & HOMMEL, B. (2007). Feature binding and affect: Emotional modulation of visuomotor integration. *Neuropsychologia*, **45**, 440-446. doi:10.1016/j.neuropsychologia.2006.06.032

FAGIOLI, S., HOMMEL, B., & SCHUBOTZ, R. I. (2007). Intentional control of attention: Action planning primes action-related stimulus dimensions. *Psychological Research*, **71**, 22-29. doi:10.1007/s00426-005-0033-3

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.

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: Evidences 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., & COLZATO, L. S. (2004). Visual attention and the temporal dynamics of feature integration. *Visual Cognition*, **11**, 483-521. doi:10.1080/13506280344000400

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.

KAAS, J. H., & HACKETT, T. A. (1999). "What" and "where" processing in auditory cortex. *Nature Neuroscience*, **2**, 1045-1047.

KAHNEMAN, D., TREISMAN, A., & GIBBS, B. J. (1992). The reviewing of object files: Object-specific integration of 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.

LEE, C. C., & WINER, J. A. (2005). Principles governing auditory cortex connections. *Cerebral Cortex*, **15**, 1804-1814. doi:10.1093/cercor/bhi057

MONDOR, T. A., HURLBURT, J., & THORNE, L. (2003). Categorizing sounds by pitch: Effects of stimulus similarity and response repetition. *Perception & Psychophysics*, **65**, 107-114.

MONDOR, T. A., & LEBOE, L. C. (2008). Stimulus and response repetition effects in the detection of sounds: Evidence of obligatory retrieval and use of a prior event. *Psychological Research*, **72**, 183-191. doi:10.1007/s00426-006-0095-x

NEUHOFF, J. G., KNIGHT, R., & WAYAND, J. (2002). Pitch change, sonification, and musical expertise: Which way is up? In R. Nakatsu & H. Kawahara (Eds.), *Proceedings of the 8th International Conference on Auditory Display*, Kyoto, Japan.

NEUHOFF, J. G., KRAMER, G., & WAYAND, J. (2002). Pitch and loudness interact in auditory displays: Can the data get lost in the map? *Journal of Experimental Psychology: Applied*, **8**, 17-25. doi:10.1037//1076 -898X8.1.17

PAVLOV, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). London: Oxford University Press.

POSNER, M. I., & COHEN, Y. A. (1984). Components of visual orienting. In: H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance: X. Control of language processes* (pp. 531-556). Hillsdale, NJ: Erlbaum.

SCHULTZ, W. (2002). Getting formal with dopamine and reward. *Neuron*, **36**, 241-263.

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. doi:10.1016/j .cogbrainres.2005.05.006

TREISMAN A. M. (1990). Variations on the theme of feature integration: Reply to Navon. *Psychological Review*, **97**, 460-463.

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

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.

ZEKI, S., & BARTELS, A. (1999). Toward a theory of visual consciousness. *Consciousness & Cognition*, **8**, 225-259.

NOTE

1. The temporal overlap scenario sketched in Figure 4 refers to the hypothetical temporal relations between coding processes in Experiment 1. These relations depend on the particular stimuli and the stimulus parameters chosen and, thus, may look very different for other stimuli, intensities, and pitch values. Thus, we do not suggest that loudness is always coded more quickly than pitch or that loudness and pitch coding or pitch and response coding always overlap in time; we suggest only that features that happen to be coded by overlapping processes are more likely to be integrated.