Project no.:	**027657**
Project full title:	**Perception, Action & Cognition through learning of Object-Action Complexes**
Project Acronym:	**PACO-PLUS**
Deliverable no.:	**D4.2.3**
Title of the deliverable:	**Integration of low-, mid-, and high-level processes**

Contractual Date of Delivery to the CEC:	31 January 2009
Actual Date of Delivery to the CEC:	19 February 2009
Organisation name of lead contractor for this deliverable:	ULg

Author(s): Justus Piater, Renaud Detry, Norbert Krüger, Florentin Wörgötter, Carme Torras, Alejandro Agostini, Juan Andrade-Cetto, Bernhard Hommel, Mark Steedman, Christopher Geib, Ronald Petrick, Tamim Asfour, Aleš Ude
Participant(s): ULg, SDU, BCCN, CSIC, UL, UEDIN, UniKarl, JSI

Work package contributing to the deliverable:	WP1, WP4.1, WP4.2, WP7.1
Nature:	R
Version:	Final
Total number of pages:	9
Start date of project:	1st Feb. 2006 **Duration:** 48 month

Project co-funded by the European Commission within the Sixth Framework Programme (2002–2006)
Dissemination Level

PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract:

This technical report provides a brief summary of the cognitive architecture developed for the PACO-PLUS project. It reviews the basic roles of each level in bottom-up (sensory-to-cognitive) and top-down (cognitive-to-motor) processing, and briefly describes how they interact. Further information is given in other Deliverables, most importantly D1.2.2, D4.1.3 and D4.3.5.

Keyword list: Architecture, Bottom-Up Processes, Top-Down Processes, Perception-Action Cycles

Table of Contents

1. Cognitive Architecture 3

2. Low Level: Online Sensorimotor Processing 4

2.1 Bottom-Up Processes 4

2.2 Top-Down Processes 4

2.3 Perception-Action Cycles 4

3. Mid Level: Memory Consolidation, Action Selection 5

3.1 Bottom-Up Processes 5

3.2 Top-Down Processes 5

3.3 Perception-Action Cycles 6

4. High Level: Reasoning, Planning, Language 6

4.1 High-level Planning 6

4.2 Rule-based Action Selection 6

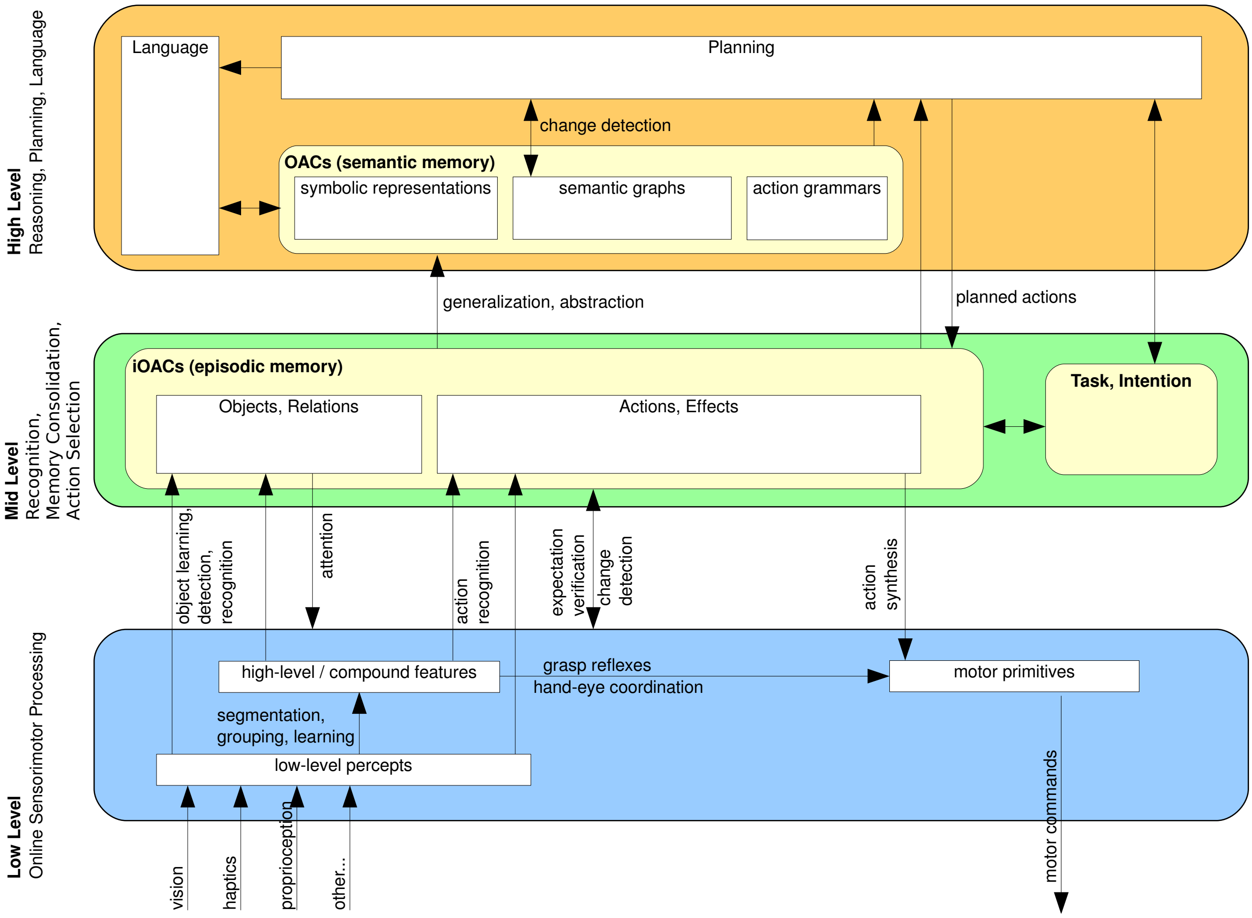


Figure 1: The PACO-PLUS Cognitive Architecture.

1. Cognitive Architecture

Figure 1 shows the PACO-PLUS system architecture [24]. It consists of three communicating processing levels. All levels are concerned with both perception and action, and serve to construct, or make use of, Object-Action Complexes (OACs). Processing generally flows clockwise in Fig. 1: Raw sensory data is received on the bottom left and is increasingly abstracted on its way up. The high level generates high-level plans based on sensory information, which are turned into concrete motor commands on their way down on the right. Executed motor commands have effects on the environment, which trigger new sensory input, closing a perception-action cycle. Each level can close perception-action cycles without going through levels above. However, the degree of adaptivity of such cycles generally increases with the number of levels involved.

The *low level* constitutes the sensorimotor interface of the robot to the physical environment. It receives raw sensor data and performs low-level processing such as feature extraction, bottom-up segmentation and grouping. The resulting digested features and low-level, bottom-up recognition results are passed up to the mid level. On the action side, desired motor behaviors are transformed into sequences of low-level motor commands for execution by the robot controllers.

The *high level* implements abstract, cognitive functions such as reasoning, planning and language. It mostly operates on symbolic representations of objects and relations, which are extracted from, and thus grounded in, concrete sensorimotor experience processed at the mid level to update its knowledge about the world. In return, actions to be performed are passed down to the mid level, again in the form of abstract, mostly symbolic representations.

The *mid level* constitutes the interface between the (mostly non-symbolic) sensorimotor representations of the low level and the (mostly symbolic) entities of the high level. Its key functional unit is an episodic memory that stores traces of sensorimotor experiences. Any regularities discovered within these traces are extracted and represented in abstract (symbolic and/or parametric rule-based) form. On the perceptual side, this yields an abstraction of raw sensory data into symbols that can be manipulated by the high level. On the motor side, these representations serve to translate symbolic action commands and their expected effects into concrete instances of effector motions and their expected sensorimotor results.

For each level, the following sections describe the links to the neighboring levels and give examples of implemented processes and their role in the multi-level system.

2. Low Level: Online Sensorimotor Processing

2.1 Bottom-Up Processes

The low level receives raw readings from perceptual and proprioceptive sensors and processes them for immediate action or for use at higher levels. Even though what is being processed (overt attention to particular objects and events) is controlled by top-down processes, the processing itself is purely reactive and does not maintain any long-term state. Elaborate examples implemented to date include:

- Feature extraction for visual detection, recognition and reconstruction (e.g., [18, 14]),
- 3D early cognitive vision (ECV): scene reconstruction in terms of patches with 5-dof pose and appearance [26, 33],
- Grasp reflexes: computation of grasp hypotheses based on relations between ECV patches [16, 15, 32],
- Computation of grasp hypotheses based on box simplifications of objects [23],
- Object pose computation and robot posture recovery [7, 6, 5]
- Oculomotor behaviors and active 3D vision [34].

2.2 Top-Down Processes

The low level receives parametrized behaviors and transforms them to motor command sequences, which are then passed down to the hardware for execution. The behaviors may be received from the mid level, or they may be generated within the low level.

2.3 Perception-Action Cycles

Some reactive behaviors have been implemented at the low level, including

- Visual servoing [16],
- Execution and verification of grasp reflexes [15, 16].

3. Mid Level: Memory Consolidation, Action Selection

3.1 Bottom-Up Processes

The mid level receives experienced interactions in terms of preprocessed perceptual input such as visual object features and object-action-effect relations from the low level [1, 2, 3], and stores them in episodic memory. From a set of experiences, new OACs can be created by generalization and abstraction (e.g. using statistical methods) and are passed to the high level. If an experience in episodic memory is the result of, or sufficiently matches, the instantiation of a given, existing OAC, then it can be used to update this OAC at the high level.

The following processes have been implemented that receive input from the low level (Section 2.1):

- Birth of objects: creation of new object representations by trying to grasp scene features and extracting features that move coherently with the gripper [25],
- Learning of 3D probabilistic object representations on the basis of ECV patches for object detection, recognition and pose estimation [19, 20, 31, 30],
- Pose estimation and grasp parameter computation using *grasp densities* [16, 15],
- Semantic scene graphs to represent relevant objects and their relations, and track them over time [15].

3.2 Top-Down Processes

The mid level receives symbolic descriptions of actions and action sequences from the high level, and passes them on to the low level in the form of parametrized behaviors. This translation process may involve parameters derived from features extracted from the current scene, and/or derived from concrete experiences stored in episodic memory.

The following processes have been implemented:

- Action synthesis based on imitation and coaching [16],
- Entropy-based action selection [13].

  Global action plans are computed at the high level of the architecture. Following the clockwise processing in Figure 1, these plans are passed to the mid level for actuation via the instantiated OAC. At this point, local decisions must be made to minimize contingencies and to maximize reward. One such method for local plan execution is to take the actions that are most informative, in the sense that they help reduce the uncertainty in the estimation of attributes [35, 8].

  This is achieved by selecting from a set of primitive actions those that maximize the predicted mutual information gain between posterior states and measurements. Maximizing the mutual information helps to avoid ill-conditioned measurements.

  The essential idea is to use mutual information as a measure of the statistical dependence between actions and attributes. The mutual information is the relative entropy between the marginal density of the attributes and the same density conditioned on the observed attribute values. When the attributes are modelled as multivariate Gaussian distributions, the parameters of the marginal density are trivially a Kalman-filter prior mean and covariance. Moreover, the parameters of the conditional density come precisely from the Kalman update equations.

Thus, in choosing a maximally mutually informative motion command, we are maximizing the difference between prior and posterior entropies [8]. In other words, we choose those actions that most reduce the uncertainty in the attributes given the observed attribute values.

3.3 Perception-Action Cycles

Several closed-loop behaviors have been implemented:

- Learning of *grasp empirical densities*. This ongoing work aims to learn empirically verified success likelihoods of object-relative grasp parameters by sampling grasp parameters from grasp hypothesis densities (Section 2.1), executing them, and monitoring their success [15],
- Pushing objects by learning about the relationship between robot movements and object responses, in order to place an object where desired, to make them graspable, or to clear access to other objects [15, 16],
- Pouring fluid content by goal-directed tilting of containers [16].

4. High Level: Reasoning, Planning, Language

The high level receives symbolic knowledge about the world and about experiences from the mid level, which it can store in a long-term, semantic (rule-based) memory and use to generate symbolic plans.

4.1 High-level Planning

The symbolic knowledge received from the lower levels forms the basis of the high-level action representation, which abstracts the capabilities of the lower levels and the working environment. This representation provides the apparatus needed to support symbolic plan generation and execution in both low-level robotic domains and high-level domains requiring language and communication.

High-level planning capabilities are supplied by the PKS planner [28, 29], a state-of-the-art knowledge-level planner. Unlike traditional planners, PKS constructs plans at the "knowledge level" by representing and reasoning about how the planner's (incomplete) knowledge state changes during plan generation. Actions are specified as STRIPS-like rules [22], in terms of their preconditions and effects. Planning is goal directed: actions are chosen by searching through the space of applicable actions, and chained together so their effects bring about a state in which the goal conditions are satisfied. PKS is able to construct conditional plans with sensing actions, and supports numerical reasoning, run-time variables [21], and features like functions that arise in real-world planning scenarios.

Details are given in Deliverable 4.3.5 [17].

4.2 Rule-based Action Selection

An alternative to the planning-based approach to action selection is to choose among available actions in a more "reactive" manner, based on their likelihood of success.

In general, actions are chosen to produce a desired change in the object given the concrete object instantiation [4, 11, 10]. The probabilities related to the success or failure in producing the changes coded in a rule are calculated from the previous experiences encoded in the OAC. The rule with highest probability of success is chosen and passed to the low levels for task execution. The outcome of the actions is fed back to the OAC in semantic memory for cause-effect generalization. That is, the action part of the OAC is updated from experience. This rule set, encoded in the high level of the architecture, is progressively refined from experience using a general to specific constructive learning, and a memory based approach [12].

The probabilities for success or failure encoded in a rule ($P+$ and $P-$ in [12]) are confidence indicators for predicting the outcome of actions since these are based on densities of samples, and not on relative frequencies, and assign to unexplored states a uniform chance of success. Therefore, rules fed only with a few successful experiences are equally valuable when evaluating their probability of success for a given task. Other performance evaluation criteria, like the m-estimate [27] or Laplace [9], would produce biased confidence measures.

Attached Papers

[1] I. B. Dutzi and B. Hommel. The microgenesis of action-effect binding. *Psychological Research*, in press.

[2] Pascal Haazebroek and Bernhard Hommel. Anticipative control of voluntary action: Towards a computational model. Lecture Notes in Artificial Intelligence. Springer, in press.

[3] S. Zmigrod, M. Spapé, and B. Hommel. Intermodal event files: Integrating features across vision, audition, taction, and action. *Psychological Research*, in press.

References

[4] A. Agostini, E. Celaya, C. Torras, and F. Wörgötter. Action rule induction from cause-effect pairs learned through robot-teacher interaction. In *Proc. International Conference on Cognitive Systems*, pages 213–218, Karlsruhe, 2008.

[5] M. Alberich, G. Alenya, J. Andrade, E. Martínez, and C. Torras. Recovering the epipolar direction from two affine views of a planar object. *Computer Vision and Image Understanding*, 112:195–209, 2008.

[6] G. Alenya and C Torras. Robot egomotion from the deformation of active contours. In S. Kolski, editor, *Mobile Robots, Perception and Navigation*, pages 1–18. pro literatur Verlag, 2007.

[7] G. Alenya and C. Torras. Monocular object pose computation with the foveal-peripheral camera of the humanoid robot Armar-III. In T. Alsinet, J. Puyol-Gruart, and C. Torras, editors, *Artificial Intelligence Research and Development*, pages 355–362. IOS Press, 2008.

[8] J. Andrade-Cetto and F. Thomas. A wire-based active tracker. *IEEE Transactions on Robotics*, 24(3):642–651, June 2008.

[9] P. Clark and R. Boswell. Rule induction with CN2: Some recent improvements. In *Proc. Fifth European Working Session on Learning*, pages 151–163, Berlin, 1991.

[10] PACO-PLUS Consortium. Scientific publications on: a) augmented OAC space-time representation, and b) coupled nonlinear estimation and control of OAC algorithms. Deliverable 7.1.1 (month 18), 2007.

[11] PACO-PLUS Consortium. Scientific publication on learning action rules on the basis of observed cause-effects and teacher instruction. Deliverable 7.2.3 (month 30), 2008.

[12] PACO-PLUS Consortium. Technical report: On-line learning of macro planning operators using probabilistic estimations of cause-effects. Deliverable 6.6 (month 30), 2008.

[13] PACO-PLUS Consortium. Scientific publication: Action selection for robotic manipulation of deformable objects. Deliverable 7.1.3 (month 36), 2009.

[14] PACO-PLUS Consortium. Scientific publication: Active building of object models. Deliverable 7.1.2 (month 36), 2009.

[15] PACO-PLUS Consortium. Technical report on generalization of affordances across objects. Deliverable 4.1.3 (month 36), 2009.

[16] PACO-PLUS Consortium. Technical report on the specification, design and implementation of the cognitive control architecture on armar. Deliverable 1.2.2 (month 36), 2009.

[17] PACO-PLUS Consortium. Technical report: Revised version of PACO-PLUS design documentation for integration of robot control and AI planning. Deliverable 4.3.5 (month 36), 2009.

[18] B. Dellen, G. Alenya, S. Foix, and C. Torras. 3D object reconstruction from swissranger sensor data using a spring-mass model. In *Proc. Intl. Conf. on Computer Vision Theory and Applications (VISAPP '09)*, February 2009.

[19] R. Detry and J. Piater. Hierarchical integration of local 3d features for probabilistic pose recovery. *Robot Manipulation: Sensing and Adapting to the Real World, 2007 (Workshop at Robotics, Science and Systems)*, 2007.

[20] Renaud Detry, Nicolas Pugeault, and Justus Piater. Probabilistic pose recovery using learned hierarchical object models. *International Cognitive Vision Workshop (Workshop at the 6th International Conference on Vision Systems)*, 2008.

[21] Oren Etzioni, Steve Hanks, Daniel Weld, Denise Draper, Neal Lesh, and Mike Williamson. An approach to planning with incomplete information. In *Proceedings of KR-92*, pages 115–125, 1992.

[22] Richard Fikes and Nils Nilsson. STRIPS: a new approach to the application of theorem proving to problem solving. *AI Journal*, 2:189–208, 1971.

[23] S. Geidenstam, K. Hübner, D. Banksell, and D. Kragic. Learning of 2D grasping strategies from box-based 3D object approximations. submitted, 2009.

[24] D. Kraft, E. Başeski, M. Popović, A. M. Batog, A. Kjær-Nielsen, N. Krüger, R. Petrick, C. Geib, N. Pugeault, M. Steedman, T. Asfour, R. Dillmann, S. Kalkan, F. Wörgötter, B. Hommel, Renaud Detry, and Justus Piater. Exploration and planning in a three level cognitive architecture. In *International Conference on Cognitive Systems (CogSys)*, 2008. (accepted).

[25] D. Kraft, N. Pugeault, E. Başeski, M. Popović, D. Kragic, S. Kalkan, F. Wörgötter, and N. Krüger. Birth of the Object: Detection of Objectness and Extraction of Object Shape through Object Action Complexes. *Special Issue on "Cognitive Humanoid Robots" of the International Journal of Humanoid Robotics*, 5:247–265, 2009.

[26] N. Krüger, M. Lappe, and F. Wörgötter. Biologically Motivated Multi-modal Processing of Visual Primitives. *The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour*, 1(5):417–428, 2004.

[27] Tom Mitchell. *Machine Learning*. McGraw-Hill, 1997.

[28] Ronald P. A. Petrick and Fahiem Bacchus. A knowledge-based approach to planning with incomplete information and sensing. In *Proceedings of AIPS-02*, pages 212–221, 2002.

[29] Ronald P. A. Petrick and Fahiem Bacchus. Extending the knowledge-based approach to planning with incomplete information and sensing. In *Proc. of ICAPS-04*, pages 2–11, 2004.

[30] Justus Piater and Renaud Detry. 3D probabilistic representations for vision and action. In *Workshop on Robotics Challenges for Machine Learning II*, 2008.

[31] Justus Piater, Fabien Scalzo, and Renaud Detry. Vision as inference in a hierarchical markov network. In *Twelfth International Conference on Cognitive and Neural Systems*, 2008.

[32] Mila Popovic, Dirk Kraft, Leon Bodenhagen, Emre Baseski, Nicolas Pugeault, Danica Kragic, and Norbert Krüger. An adaptive strategy for grasping unknown objects based on co-planarity and colour information. to be submitted.

[33] N. Pugeault. *Early Cognitive Vision: Feedback Mechanisms for the Disambiguation of Early Visual Representation*. Vdm Verlag Dr. Müller, 2008.

[34] Aleš Ude and Tamim Asfour. Control and recognition on a humanoid head with cameras having different field of view. In *Proc. IAPR Conf. Pattern Recognition (ICPR)*, Tampa, Florida, December 2008.

[35] T. Vidal-Calleja, A.J. Davison, J. Andrade-Cetto, and D. W. Murray. Active control for single camera SLAM. In *Proc. IEEE International Conference on Robotics and Automation*, pages 1930–1936, Orlando, May 2006.

Psychological Research
DOI 10.1007/s00426-008-0161-7

ORIGINAL ARTICLE

The microgenesis of action-effect binding

Ilona B. Dutzi · Bernhard Hommel

Received: 5 July 2007 / Accepted: 29 August 2008

Abstract Ideomotor theories of human action control assume that performing a movement leads to the automatic integration of the underlying motor pattern with codes of its perceptual consequences. We studied the microgenesis of action-effect integration by varying the mapping of action effects upon actions from trial to trial. Experiments 1 and 2 showed that perceiving a tone repetition systematically affects one's tendency to carry out the response that produced that tone in the previous trial, suggesting that even the unintentional production of a stimulus creates a temporary binding of that stimulus with the action that brought it about. Experiments 3 and 4 extended this finding in suggesting that the integration and/or retrieval of action effects is modulated by attentional factors: Ongoing performance is more impacted by action effects if they are salient or match the current attentional set.

Introduction

People plan and perform voluntary actions in order to reach particular, intended goals, that is, to modify particular states of affairs or create particular events. Obviously, they can do so only if they have reliable knowledge at their disposal regarding which kind of action is likely to create the intended event. According to ideomotor theories of voluntary action (James, 1890; Lotze, 1852) and, to some degree, Piaget's (1946) sensorimotor approach to cognition, this knowledge is acquired "on the fly": Carrying out movements is assumed to be accompanied by a more or less automatic process of self-perception that integrates, without much ado, the motor patterns underlying the movement with the codes of that movement's perceptual consequences. In other words, actions become automatically associated with codes of their perceivable effects. This bilateral association provides the individual with a retrieval cue that allows creating that effect intentionally: One only needs to "think of" or "anticipate" (i.e., internally activate the codes of) particular action effects in order to prime and activate the action that has been experienced to produce that effect before.

Although this issue was neglected for quite some time, numerous recent studies provide increasing evidence that action effects are indeed picked up in an automatic fashion (for an overview, see Hommel & Elsner, 2009): People quickly acquire bilateral associations between actions and novel effects, such as keypress-contingent tones of a particular pitch or lights of a particular location, whether these effects are relevant to, or useful for the task at hand (Hoffmann, Sebald, & Stöcker, 2001; Hommel, 1993; Ziessler, 1998) or not (Beckers, De Houwer, & Eelen, 2002; Elsner & Hommel, 2001; Hommel, 1996). As studies using PET and fMRI have shown, once an action effect has been acquired its mere perception primes apparently associated motor structures (in the caudal supplementary motor area; Elsner et al., 2002; Melcher, Weidema, Eenshuistra, Hommel, & Gruber, 2008).

I. B. Dutzi
Bethanien Hospital, Geriatric Center, Heidelberg, Germany

B. Hommel
Leiden University Institute for Psychological Research and Leiden Institute for Brain and Cognition, Leiden, The Netherlands

B. Hommel (✉)
Department of Psychology, Cognitive Psychology Unit,
University of Leiden, Wassenaarseweg 52,
2333 AK Leiden, The Netherlands
e-mail: hommel@fsw.leidenuniv.nl

Even though the bulk of the evidence suggests that action-effect learning occurs spontaneously and without any intention to learn, Ziessler, Nattkemper, and Frensch (2004) have argued that effective action-effect acquisition depends on the active anticipation of the effects and is thus under attentional control. In their study, participants carried out pairs of manual responses signaled by visual letters ($S_1 \rightarrow R_1$, $S_2 \rightarrow R_2$). The second stimulus was systematically related to the preceding response so to allow for acquiring R_1–S_2 associations—which Ziessler et al. consider comparable to action-effect associations. As evidence for R_1–S_2 learning was obtained under undistracted conditions but not when participants were in addition to the sequential task to count tones presented in the R_1–S_2 interval, the authors conclude that R_1–S_2 acquisition cannot be automatic. But this conclusion is neither obvious nor necessary. First, ideomotor approaches claim that action-effect learning is automatic in the sense of not requiring an intention to learn, but they do not speak to the amount of cognitive resources involved. For instance, it is not unreasonable to assume that action-effect bindings need to be consolidated in order to affect subsequent behavior. As memory consolidation is known to be resource demanding and fragile (Jolicœur & Dell'Acqua, 1998), it may well be that it suffers from an overlapping task, such as tone counting. Second, even though it does not involve overt motor output, counting a tone may well be considered an intentional action. This means that Ziessler et al.'s tone-counting condition turned the original R_1–S_2 sequence into one where a third action intervened between R_1 and S_2, rendering it a R_1–R_2–S_2 sequence. If so, people might well have acquired R_2–S_2 associations, but that they failed to acquire R_1–S_2 associations is hardly surprising. Finally, the group that eventually showed the largest R_1–S_2 learning effects also showed by far the best performance on all measures from the very first trials on. For instance, their average reaction time for the first 12 performances of R_1 (the response that preceded and could thus not be affected by the tone) was already about 100 ms faster than the average of any of the other three groups. This strongly suggests major differences in motivation, which may also account for more efficient learning. In sum, we doubt that the available evidence provides strong support for a selective integration mechanism. On the contrary, numerous findings support the ideomotor expectation that carrying out a movement is indeed accompanied by the automatic (i.e., unintentional) integration of its perceptual consequences.

The present study focused on the microgenesis of this integration process, that is, the emergence of individual action-effect associations. According to the Theory of Event Coding (TEC) of Hommel, Müsseler, Aschersleben, and Prinz (2001) stimulus and action events are integrated in two phases. The first, *activation phase* consists of activating codes of a particular stimulus and/or action feature, be that internally driven, as in the case of action planning (accomplished by "anticipating" the intended action's attributes), or externally driven, such as when a stimulus event is perceived. The second, *integration phase* serves to bind the activated features together, hence, to integrate them into a sort of event file (Hommel, 1998). These event bindings may be actively maintained, such as when an action plan is held in preparation (Stoet & Hommel, 1999), or decay over time. In any case, however, event bindings seem to survive 1 s or longer (Hommel, 1998; Hommel & Colzato, 2004).

Here we applied TEC integration logic to action-effect integration. TEC claims that if the activations of codes (be they stimulus- or action-related) overlap in time, they get integrated. Hence, if the codes of an action plan are still activated to some degree when the effects of that action are coded, action and effects should become part of the same representational structure. Given that the codes of action plans commonly show activation 250 ms or longer after the corresponding action is carried out (Stoet & Hommel, 1999; Hommel, 1994), there are reasons to believe that the overlap is sufficient at least for immediate effects triggered by the action's onset. Indeed, studies of long-term action-effect acquisition have shown that actions and effects are spontaneously associated if the effects follow the action onset by up to 1 s but not longer (Elsner & Hommel, 2004)—at least if the action-effect interval is not "bridged" by intervening events (cf., Reed, 1999). Likewise, if participants estimate the extent their actions have caused a particular event, the accuracy of their judgments decreases considerably if actions and effects are separated by more than about 2 s (Shanks, Pearson & Dickinson, 1989). With respect to the short-term binding of stimuli and responses, it has been shown that stimuli are integrated with responses if they appear in a temporal neighborhood of about half a second but not if they are separated from the response by about 2.5 s (Hommel, 2005).

If actions and effects are spontaneously (i.e., non-intentionally) integrated into action-effect bindings and if these bindings have a lifetime beyond the presentation of the effect, the way they are bound together should affect subsequent performance. Assume, for instance, a left-hand keypress is heard to produce a low-pitched tone, in a task where high and low tones can appear and left and right keypresses are carried out. If the co-occurrence of left-hand keypress and low-pitched tone creates a binding between the codes LOW and LEFT, presenting a high or low tone shortly thereafter (i.e., while the binding is still intact) should systematically bias response selection. Figure 1 shows how. Given the high and low tones are the perceptual alternatives in the present context, the participant is likely to represent these two possibilities as

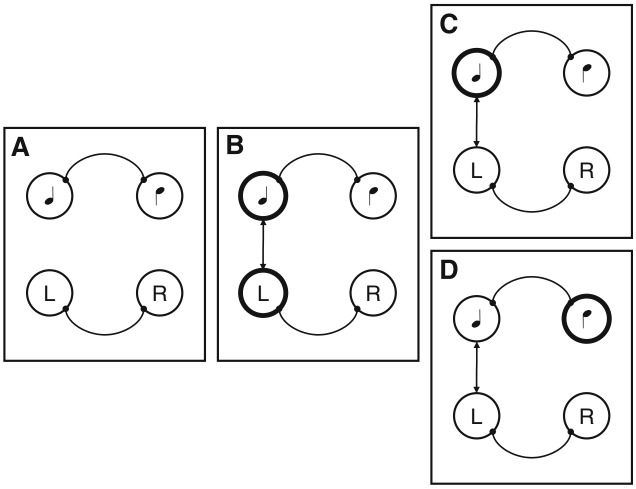


Fig. 1 Illustration of the creation and retrieval of action-effect bindings. **a** Being exposed to high- and low-pitched tones leads to the cognitive representation of these tones (low and high note for low and high tones, respectively), which given that the two tones are alternatives in the present context are connected by a mutually inhibitory link. Likewise, carrying out left and right responses leads to the representation of these (again mutually exclusive) alternatives (*L* and *R* for left and right responses, respectively). **b** Carrying out a left response followed by a low tone leads to the activation of the corresponding codes, which again leads to their integration (indicated by the *double arrow* between them). For the lifetime of the binding, the two codes to act as an unit. **c** Subsequently perceiving another low tone reactivates the corresponding code, which spreads activation to the left response code it is still integrated with. That is, a stimulus repetition primes a response repetition by biasing the competition between response codes toward the left code. **d** Subsequently perceiving the stimulus alternative (a high tone) activates the corresponding code, which will inhibit the code of the stimulus alternative (the low tone) via the inhibitory link. Given that the low tone is still integrated with the left response, this inhibition will spread to the left response code. Consequently, the competition between response codes is biased against the left code, so that stimulus alternation facilitates response alternation

sketched in Fig. 1a, where the codes for high and low tones are connected by an inhibitory link (see Bogacz, 2007). The same logic applies to the two alternative responses (the left and right keypress or L and R), which are also shown in this figure. If we assume that tones and responses vary independently and are thus uncorrelated, there are no long-term associations between tones and responses. However, according to our reasoning, a single co-occurrence of low tone and left response should induce a binding between their representations, as indicated in Fig. 1b.

What would happen if the tone repeats? As shown in Fig. 1c, activating the code of the low tone should prime the still bound response, the left keypress that is. This means that a stimulus repetition should induce a tendency to repeat the response as well. Now consider what a tone alternation would imply. As shown in Fig. 1d, presenting a high tone would activate the corresponding code, which is not bound to any response (if we ignore previous trials for a moment). However, given the inhibitory link between the two tone representations, activating the code of the high tone should lead to the inhibition of the low-tone code. Given that this code is still bound with the left response code, inhibition will spread to that code as well. This follows from the integrated competition hypothesis suggested by Duncan and colleagues (Duncan, 1996; Duncan, Humphreys & Ward, 1997). They pointed out that the distributed cortical representation of perceptual and action codes calls for integration mechanisms that create coherent object-action compounds. Members of such a compound benefit from competitive gains achieved by other members of the same compound, so that, say, integrating RED with ROUND when processing the image of a cherry has the consequence that increasing the activation of the RED code also supports the ROUND code in its competition with other shape-related codes. The flipside of integrated competition is that losses in the competition also spread among members, so that outcompeting the RED code when seeing a banana somewhat later will also weaken the ROUND code associated with it. In other words, integrated elements win together and lose together. Applied to our example, this means that binding LOW and LEFT weakens LEFT if LOW loses against HIGH. Given that left and right responses are the only alternatives, this again implies that perceiving a high tone would bias response selection toward the right response, which would benefit from the indirect inhibition of the left response code.

Available evidence from stimulus-response integration studies provides support for both implications. For one, repeating stimulus features have been shown to speed up response repetitions as compared to response alternations (Hommel, 1998; Hommel & Colzato, 2004), suggesting that stimulus repetition indeed induces a response-repetition tendency. For another, alternations of stimulus features have been observed to speed up response alternations, sometimes even more than stimulus repetitions speed up response repetitions (e.g., Hommel & Colzato, 2004). Along the same lines, with multidimensional stimuli, response repetitions are particularly (i.e., over-additively) fast if signaled by a stimulus that repeats all the features of the previous stimulus (Bertelson, 1963), whereas response alternations are particularly slow under these circumstances (Hommel, Memelink, Colzato, & Zmigrod, 2008). Hence, stimulus alternations indeed seem to induce a response-alternation tendency.

According to these considerations perceiving a tone that does or does not match a just-experienced response-produced tone should systematically bias the decision to perform a left or right keypress. Importantly, this should be the case independently of previous experiences, hence,

Table 1 Conditions in Experiments 1 and 2

Condition	Induction S → R_1 → E_A	Test E'_A → R_2
congruent	***** ↗ R_1 → ♩ ***** ↘ R_1 → ♪	♩ → ? ♪ → ?
incongruent	***** ↗ R_1 → ♪ ***** ↘ R_1 → ♩	♩ → ? ♪ → ?
no-go	***** ↗ R_1 → ♪ ***** ↘ R_1 → ♩	

Each trial consisted of two parts. In the induction part, a row of asterisks (Stimulus) triggered a free-choice response (R_1), which was followed by a randomly chosen high or low pitched tone (auditory action effect E_A). In congruent or incongruent go trials of the test part, a high or low pitched tone (which was congruent or incongruent with the preceding action effect) triggered another free-choice response (R_2)—the dependent measure (?) being the type of response (same as vs. different from R_1)

even if the overall probabilities for a high and low tone to follow a left or right response are equal. We tested this prediction as sketched in Table 1. Participants carried out free-choice responses by pressing a left or right key (for a discussion and validation of this technique, see Elsner & Hommel, 2001; Hommel, 2007). Each trial consisted of two parts. In the first, *induction* part participants made a freely chosen response (R_1) to a non-discriminative visual trigger stimulus (S). This response produced one of two auditory effects (E_A), a low- or a high-pitched tone. Importantly, the mapping of response keys to pitch varied randomly from trial to trial, so to prevent any incremental response-effect learning across the experimental session. One-second later, in the *test* part of each trial, participants encountered one of the two effect stimuli (E'_A), which now served as go signal (in 75% of the trials) to perform another freely chosen response (R_2). The measure of interest was the response choice in the test part (i.e., R_2). In particular, we analyzed the tendency to repeat the previous response ($R_2 = R_1$) as a function of the relationship between the effect tone E_A and the go-signal tone E'_A. According to our hypothesis, participants should be more likely to repeat a response if the two tones match ($E'_A = E_A$) than if the tones do not match ($E'_A \neq E_A$), because the tone's code should still be bound with the response that just had produced it.

Experiment 1

Experiment 1 was conducted as a first test whether action-related codes are spontaneously integrated with codes of their effects, as suggested by TEC. If so, we would expect response-repetition rates (%RR) to be higher if the R_2-go signal (E'_A) matches the preceding action effect (E_A) in pitch than if it does not.

Method

Participants

Twenty students served as paid participants. As was the case for all participants of this study, they reported having normal or corrected-to-normal vision and audition and were not familiar with the purpose of the experiment.

Apparatus and stimuli

Visual stimuli (a row of 13 white-on-black asterisks) were presented on a computer monitor and auditory stimuli (sinusoidal tones of 400 and 800 Hz) through external loudspeakers to the monitor's left and right. Responses were made by pressing the left or right of two external microswitches with the corresponding index finger. The experiment was controlled by a standard PC running under ERTS (Beringer, 1994).

Procedure

Each trial consisted of an *induction* part, to induce a particular action-effect binding, and a *test* part, to diagnose the presence of such bindings. Table 1 shows the sequence of events. After an intertrial interval of 3,000 ms, the asterisk string (S) appeared for 300 ms, requesting a speeded left or right keypress (R_1). Participants were instructed to choose the key randomly and to avoid any strategy apart from using the keys about equally often. If a response was made a randomly selected effect tone (E_A) was presented for 100 ms, its onset being synchronized with the keypress. Due to the random selection procedure, keypresses and tone pitches were uncorrelated, that is, in a given trial each keypress had the same probability to trigger either a low or high tone. Participants were told that these tones were completely irrelevant for the task and that there would be no systematic relationship between keypress and pitch.

In the second, test part of each trial one of the two effect tones was used as go signal (E'_A) to signal a second free-choice reaction (R_2) in 75% of the trials; in the remaining 25% no tone appeared and no second response was to be performed (no-go trials). No-go trials were used to work against some of the most obvious strategies in free-choice tasks, such as choosing responses according to a standard predetermined pattern. In go trials one of the two tones sounded for 100, 1,000 ms after the previous effect tone had been presented. In 50% of these go trials the tone was the same as the previous effect tone (congruent trial); in the other 50% the signal tone was the alternative tone (incongruent trial). Participants were instructed to respond to the tone as quickly and as spontaneously as possible by pressing a randomly chosen response key and to refrain from responding if no second tone would occur. It was emphasized that only the *presence* of a tone mattered for the execution of R_2 while its *pitch* would be neither relevant nor informative. Participants were also urged to use both keys and not to apply any strategy. The program waited up to 1,500 ms for a response. Responses with reaction times exceeding 1,500 ms were counted as missing, those faster than 100 ms as anticipation, and responses in no-go trials as false alarms. All these errors were fed back to the participants. Following ten randomly drawn practice trials three blocks of 64 randomly ordered trials each were administered. After the session participants were asked whether they had obeyed to the instruction and had guessed the purpose of the experiment.

Results and discussion

Our dependent measure of choice is very sensitive to individual strategies, which may conceal or even prevent the possible impact of go stimuli on response choices. Particularly damaging would be strategies that determine response choices long before the go stimulus is presented, so that the selection process we intended to bias is already completed. Accordingly, we not only took measures to work against some of the strategies by speeding response selection and including no-go trials, but we also excluded participants that were likely to apply a particular "pre-selection" strategy. For this reason, we only considered participants who produced less than 20% false alarms and at least 90% correct trials altogether, and who did not report having used a response rule. All participants passed these criteria and no-one reported having paid any attention to pitch or having guessed the purpose of the experiment. In fact, most of them believed that reaction time was the important dependent variable. We also excluded participants if their mean %RR was lower than 10% or higher than 90%, which we consider strong evidence of an alternation or repetition strategy, respectively. This applied to two participants. After excluding trials with response omission (0.3%) or anticipation (0.4%) individual %RRs were calculated as a function of congruency (see Table 1 for the coding scheme).[1]

In the induction part of the trials the two keys were pressed equally often and their frequencies (48.5 vs. 51.5%) did not differ from chance. This observation, which we also made in the following experiments, confirms that participants experienced all possible response-effect couplings about equally often. The mean %RR in the test part was 39.1%, but the repetition rate was modulated by $E_A - E'_A$ congruency: As shown in Table 2, congruent trials produced more response repetitions than incongruent trials, $t(17) = 4.86$, $p < 0.01$. That is, as expected, stimulus repetitions were associated with more response repetitions, suggesting that the present response choice was affected by the relationship between the previous response and its auditory effect.

It is interesting to note that the response-repetition frequencies for congruent and incongruent conditions were not distributed evenly around 50% but shifted toward response alternations (i.e., around 39.1%). There are at least two possible accounts for this observation, which we will also make in the following experiments. The first account considers that people are often biased toward response alternations, as can be seen in faster reaction times with response alternations that repetitions, presumably reflecting a general misconception about statistical probability (Bertelson, 1961; Soetens, Boer, & Hueting, 1985)—also known as gambler's fallacy. Interestingly, response alternations were faster than repetitions (335 vs. 357 ms) in the present experiment as well, $t(17) = 3.01$, $p < 0.01$. Hence, even though our study does not provide a "pure" measure of the alternation bias, the fact that it has been so often observed in other studies may be taken to suggest that our participants also showed such a bias. This again might suggest that our congruent and incongruent conditions were indeed symmetrically distributed around a mean that would

[1] We did not consider reaction times (with one exception in another context below) because, as demonstrated and discussed by Elsner and Hommel (2001), it is impossible to predict and interpret their pattern in the present free-choice task. For instance, fast responses in the congruent condition may indicate that (a) the response was particularly spontaneous, and thus free from strategic considerations, suggesting that conditions were particularly good for tones to affect response selection; or that (b) the response was preplanned already and thus under full strategic control, suggesting that conditions were particularly bad for tones to affect response selection. Slow responses in the congruent condition may be interpreted to reflect (a) particularly strong contributions from strategic considerations (a lot of thinking before responding), which one would expect to minimize the impact of the tone; but they may just as well reflect (b) a delay due to extended competition between strategic top-down factors and tone-driven bottom up factors, which again would point to a particularly strong contribution from the tone. Multiple interpretations also exist for fast or slow responses in the incongruent condition, all suggesting that reaction times are a misleading measure in the present context.

Table 2 Mean response repetition frequencies (in %) and response-repetition biases (congruent–incongruent) for Experiments 1–4 as a function of $E - E'$ congruency (match between effect of R_1 and go signal for R_2), action-effect modality, and modality of the task-relevant go signal for R_2

Auditory action effect (pitch)			Visual action effect (color)		
Congruent	Incongruent	Bias	Congruent	Incongruent	Bias
Experiment 1					
Auditory go signal					
43.5 (12.1)	34.6 (12.1)	8.9**			
Experiment 2					
Auditory go signal					
49.1 (12.3)	41.9 (15.0)	7.1**			
Experiment 3					
Auditory go signal					
47.6 (22.1)	40.5 (19.9)	7.1**			
Visual go signal					
50.1 (21.6)	46.9 (18.4)	3.2**			
Experiment 4					
Auditory go signal					
46.4 (14.3)	41.4 (13.6)	5.0**	42.9 (12.7)	44.9 (14.4)	−2.0 NS
Visual go signal					
47.1 (19.9)	41.9 (17.6)	5.2**	46.3 (19.1)	42.7 (18.3)	3.6**

Standard deviations of means are given in parentheses

NS non-significant bias

** Significant bias, $p < 0.01$

ideally approach 50%, but the whole distribution was torn to the lower half because of the gambler's fallacy.

The second account holds that our participants were not biased toward repetition or alternation in principle, and that the outcome for the congruent condition represents something like a neutral baseline. Indeed, it will turn out that the 43.5% we observed in the congruent condition of Experiment 1 is the lowest estimate of the present study, and that the other experiments will produce estimates very close to 50%. If so, the main impact of stimulus repetitions and alternations would consist in stimulus alternations biasing people toward response alternations. In other words, the effect sketched in Fig. 1d would be much stronger than the one in Fig. 1c. As mentioned earlier, this would fit with occasional observations that stimulus-response alternations produce faster and more accurate responses than complete stimulus-response repetitions, at least numerically (e.g., Colzato, Fagioli, Erasmus, & Hommel, 2005; Colzato, van Wouwe, & Hommel, 2007; Hommel & Colzato, 2004). Indeed, given that repetition-induced priming of previous bindings and alternation-induced integrated competition are different types of processes, there is no reason to believe that the reaction-time benefits they produce should be of exactly the same size.

As we neither have a pure measure of possible general alternation biases nor a noise-free measure of binding reactivation and integrated competition, it is premature to try deciding between these two interpretations. Importantly, however, they both rest on the same assumption, namely, that perceiving a self-produced stimulus event creates a temporary binding of the codes underlying the action and the codes representing the perceived event. As a consequence, perceiving the same event or its alternative systematically biases response selection. Taken altogether, Experiment 1 provides first evidence for our hypothesis that a single pairing of an action and an effect is sufficient to integrate their cognitive representations, and that this integration has a systematic effect on subsequent response selection.

Experiment 2

Even though the outcome of Experiment 1 is consistent with our expectation that action-effect binding affects subsequent response-selection processes, there is an alternative interpretation. Our participants had the task of producing random responses and response sequences, which is known to be very hard to do. One way to make this task easier and to still meet the task requirement of getting close to a 50:50 distribution of response repetitions and alternations would be to strategically repeat the response whenever the stimulus repeats. Note that this strategy only works well if the probability of stimulus repetition versus alternation in go trials is also 50:50. If this ratio would be drastically changed, such as if stimulus alternations would be much more frequent than stimulus repetitions, such a matching strategy would be bound to fail: either response alternations would now also become much more frequent than response repetitions or participants would notice that a matching strategy makes little sense and simply no longer apply it. This was the logic underlying Experiment 2, which replicated Experiment 1 with a 25:75 probability of stimulus repetitions and alternations. According to a strategic interpretation of the congruency effect, this manipulation should eliminate the effect, whereas an interpretation in terms of action-effect binding predicts the same outcome as in Experiment 1.

Method

Twenty-one students served as paid participants. The method was exactly as in Experiment 1 with only one exception: the go trials of the test phase did not consist of 50% congruent and 50% incongruent conditions but of 25% congruent and 75% incongruent conditions; i.e., the trigger tone matched the previous action effect tone in only one quarter of the trials.

Results and discussion

Applying the same criteria as in Experiment 1 led to the exclusion of one participant. Again, trials with response omissions (<0.9%) and anticipations (<0.3%) were excluded. The overall response-repetition rate in the test part was 43.7%, which is higher than in Experiment 1 and statistically no longer different from chance. Clearly, this observation does not support the idea that participants might have strategically matched the response repetition rate to the stimulus repetition rate. The response repetition rate was again modulated by $E_A - E'_A$ congruency, $t(19) = 3.07$, $p < 0.01$, due to that congruent trials produced more response repetitions than incongruent trials did (see Table 2). An ANOVA on the combined data from Experiments 1 and 2 did not yield any hint to an interaction between experiment and congruency effect, $p > 0.5$, confirming that the congruency effect was equivalent in the two experiments. Taken together, these findings suggest that the congruency effect does not reflect a deliberate response selection strategy but rather represents an automatic by-product of action-effect binding.

Experiment 3

Our next experiment was conducted to see how automatic the impact of action-effect bindings really is and whether, or to what degree it is sensitive to attentional effects, that is, to the task relevance of the effect's perceptual characteristics. In Experiments 1 and 2, tones were used as both action effects (E_A) and R$_2$-go signals (E'_A). Accordingly, although neither the pitch nor the presence of the action effect was of any relevance, tones did play an important role and could not be ignored entirely. It may have been this, somewhat indirect type of task relevance that drew sufficient attention to the effects to integrate and bind them with the responses and/or to retrieve the just-bound response when the effect stimulus was encountered again.[2] If so, it should be possible to reduce or eliminate the impact of action-effect bindings on ongoing response selection by defining the R$_2$-go signal in another than the auditory modality, so that tones are no longer of any relevance. This is what we did in Experiment 3. In an auditory-go condition we replicated Experiment 1 by using again an auditory E'. But we also ran a visual-go condition, where the R$_2$-go signal was a visual stimulus (E'_V). Although no longer of any relevance for the task, the tone was still presented as E'_A, thus accompanying the visual go signal in go trials and as the only stimulus in the test part of no-go trials. If task relevance affected the creation and/or retrieval of action-effect bindings we would expect the response-rate effect—that is, higher response-repetition rates if the R$_2$-go signal matches the preceding action effect—in the auditory-go condition but not (or less so) in the visual-go condition.

Method

Another 26 female and 19 male students were randomly assigned to two groups of 23 and 22 participants, respectively. For the first, auditory-go group the method was exactly as in Experiment 1. For the second, visual-go group several modifications were introduced. The relevant go signal in the test part of each trial was not a tone but a red 3x3 cm square (E'_V)[3] appearing for 300 ms at screen center. Just like the tone in the auditory-go group, the square was presented in 75% of the trials to signal R$_2$ (go trials) and participants were to withhold R$_2$ in the remaining no-go trials. The pitch of the tone matched the previous action-effect tone in 50% of the go trials and the alternative tone in the other 50%. It was pointed out to the participants that both presence and pitch of the tone would be completely irrelevant for the task.

[2] Note that the design of our study (or of any other analysis of sequential effects) does not allow disentangling possible effects on the creation of action-effect bindings (i.e., integration) and on their retrieval. Obviously, bindings can only be retrieved if they have been created earlier, so that changes in the impact of action-effect bindings on performance may be due to changes in the binding process, changes in the retrieval process, or both. We will get back to this issue in "General discussion".

[3] In keeping with the terminology introduced in Experiment 1 we refer to the visual go signal as E'_V. However, note that in the induction part of Experiment 3 there were only auditory action effects (E_A) but no visual action effects (E_V).

Results and discussion

Applying the same criteria as in Experiment 1 led to the exclusion of three members of the auditory and of two of the visual group. Again, trials with response omissions (<0.6%) and anticipations (<0.1%) were excluded. The overall response-repetition rate in the test part was 46.3%, which is almost the same as in Experiment 2 and statistically not different from chance. Mean %RR were analyzed as a function of $E_A - E'_A$ congruency (i.e., whether the two tones in each trial matched or not) and R_2-go-signal modality (i.e., whether R_2 was carried out in response to the second tone or a color square; see Table 2). The only reliable finding was a main effect of tone congruency, $F(1,38) = 8.69$, $p < 0.005$, while the interaction with go-signal modality was far from significance, $F(1,38) < 1$. That is, irrespective of the tone's task relevance, R_1 is repeated more often if the tone in the test phase matches the action-effect tone.

All in all, the outcome of Experiment 3 is somewhat mixed. Statistically speaking, task relevance had no impact on the congruency effect, suggesting that the auditory action effects were integrated and retrieved in either go-signal condition. However, numerically the induced bias in the visual condition was not even half as big as that obtained in the auditory condition. Moreover, auditory stimuli and their impact on perceptual processing have been demonstrated to be more salient, hence, to rely much less on attention than visual stimuli (Posner, Nissen, & Klein, 1976). One therefore may argue that Experiment 3 provides a rather conservative test of the impact of attention.

Experiment 4

To provide a more sensitive test we ran Experiment 4, where responses produced both auditory and visual effects. We also presented stimuli of both modalities in the test part of the go trials and varied their relevance. In one block, R_2-go signals were auditory, just like in Experiments 1–3, which rendered the visual stimuli in either part of the trial irrelevant. The saliency hypothesis suggests that auditory action effects should be integrated under such conditions while visual effect may not. If so, repeating tone pitch ($E'_A = E_A$) should lead to higher response-repetition rates than alternating pitch, whereas repeating color ($E'_V = E_V$) should yield the same rates as alternating color. In another block, R_2-go signals were visually defined, which rendered the auditory stimuli in either part of the trial irrelevant. According to the saliency hypothesis and in view of Experiment 3, we would expect that both auditory and visual action effects are integrated and retrieved, so that response-repetition rates should depend on whether pitch or color is repeated.

Method

Another 27 female and 23 male students served as paid volunteers. The method was as in Experiment 3 (visual-go group) with the following exceptions. With regard to the induction part, performing R_1 now caused the simultaneous presentation of a low- or high-pitched tone (for 100 ms) *and* a red or green square at screen center (for 200 ms); hence, each R_1 had both an auditory and a visual effect.

In the test part of the trials, three independent variables were manipulated: the modality of the R_2-go signal (tone or square), the congruency between the pitch of the action-effect tone from the induction part (E_A) and the pitch of the tone presented in the test part (E'_A), and the congruency between the color of the action-effect square from the induction part (E_V) and the color of the square presented in the test part (E'_V). Like in the visual-go group of Experiment 3, there were two stimuli in the test part of go trials, a low- or high-pitched tone and a red or green square. However, in a given block only one of them was task-relevant by virtue of signaling a go trial, whereas the other was entirely irrelevant.

The experimental session consisted of two blocks, an auditory-go block, where no-go trials were defined by the absence of a tone in the test part of the trial, and a visual-go block, where no-go trials were defined by the absence of a square in the test part of the trial. Block order was balanced across participants. Each block was composed of 10 randomly drawn practice trials and 192 randomly ordered experimental trials. The 192 experimental trials comprised 144 go and 48 no-go trials, so that the go probability was again 75%. The 144 go trials were composed of 36 trials in which both the tone and the color presented in the test part matched the action effects of the preceding induction part (i.e., $E'_A = E_A$ and $E'_V = E_V$), 36 trials in which only the auditory stimuli matched (i.e., $E'_A = E_A$ and $E'_V \neq E_V$), 36 trials in which only the visual stimuli matched (i.e., $E'_A \neq E_A$ and $E'_V = E_V$), and 36 trials in which neither the auditory nor the visual stimuli matched (i.e., $E'_A \neq E_A$ and $E'_V \neq E_V$). To ensure that participants registered the visual action effects even in the auditory-go block, the instruction emphasized that they should always fixate the center of the computer screen.

Results and discussion

Applying the same criteria as in Experiment 1 led to the exclusion of 10 participants. Trials with response omissions (<0.2%) and anticipations (<0.1%) were excluded. Mean %RR were calculated for each participant as a function of auditory ($E'_A - E_A$) congruency, visual ($E'_V - E_V$) congruency, and modality of the relevant go signal (see Table 2). A corresponding 2 × 2 × 2 ANOVA produced two significant results: a main effect for auditory congruency, $F(1,39) = 17.98$, $p < 0.001$, and an interaction of visual congruency and go-signal modality, $F(1,39) = 9.26$, $p < 0.005$. As Table 2 shows, congruent pitch yielded a higher rate of response repetitions independently of go-signal modality whereas congruent color affected the repetition rate only if go trials were defined by the presence or absence of visual stimuli. Indeed, separate t-tests revealed a highly significant effect of color congruency in the visual-go block, $t(39) = 3.27$, $p < 0.005$, but not in the auditory-go block, $t(39) = 1.39$, $p > 0.05$. This pattern supports an account in terms of stimulus saliency: Action effects are integrated and retrieved if they are either relevant to the task or salient enough to attract attention in a bottom-up fashion.

Another important finding of Experiment 4 is that it for the first time demonstrates the integration of multiple action effects. Although previous studies on action-effect acquisition employed a variety of to-be-learned effect stimuli they were always restricted to one type of stimulus at a time. Yet, both auditory and visual action effects influenced performance in the visual-go condition of Experiment 4, which suggests that participants had integrated their actions with both pitch and color.[4]

General discussion

Our study tested a TEC approach to the integration of actions and their effects. We hypothesized that the likely temporal overlap of activation of action- and effect-related codes induces the temporary binding of those codes. This binding may affect subsequent behavior by biasing it toward response repetition in case of a stimulus repetition, biasing it toward response alternation in case of stimulus alternation, or both. Consistent with this expectation, Experiment 1 showed that varying the pitch of a go signal systematically affects the tendency to repeat or alternate the response that was just experienced to produce a tone of that pitch. Together with the outcome of Experiment 2, which rules out a strategic interpretation of the response repetition bias, this suggests that codes of that action are still bound with codes of the tone it produced. As a consequence, re-activating the tone-related code spread activation to the corresponding action-related code, thus priming the previous action as indicated in Fig. 1c, while activating the alternative tone code led to the inhibition of the codes of both the previous tone and the previous response, resulting in a preference for response alternation (Fig. 1d). Interestingly, the distribution of response-repetition frequencies was shifted toward response alternation in all experiments. This might reflect a general impact of the gamblers fallacy and represent the same bias that has been shown in studies of sequential stimulus and response effects (Bertelson, 1961; Soetens et al., 1985). Alternatively, it might indicate that alternations of action-effect stimuli bias subsequent response selection more toward response alternations than effect repetitions bias selection toward response repetitions. In other words, the priming of response repetitions as sketched in Fig. 1c may be less efficient than the inhibition of response repetitions as sketched in Fig. 1d. The present study does not allow disentangling these two possibilities, which calls for a more detailed experimental analysis. However, both possibilities imply that actions and effects are spontaneously integrated into temporary bindings, which supports our main hypothesis.

As the action effect in Experiment 1 was not relevant or informative, effect integration seems to be spontaneous in the sense that it does not require the explicit intention to learn about those effects. This supports Elsner and Hommel's (2001) assumption that effect integration is an automatic by-product of moving and acting. However, this does not mean that goals and intentions, and the attentional set they bring about, have no impact on effect integration and/or retrieval (Hommel et al., 2001). To the contrary, Experiments 3 and 4 provide evidence that the likelihood with which action-effect bindings affect performance depends on both bottom-up and top-down attentional factors. If an effect is salient enough to attract attention in a bottom-up fashion, as can be assumed for tones (Posner et al., 1976), action effects impact behavior even if they are neither directly nor indirectly related to the task at hand. This fits well with observations from learning studies, where auditory (e.g., Hoffmann et al., 2001; Hommel, 1996) and electrocutaneous (Beckers et al., 2002) action effects were spontaneously acquired and retrieved in otherwise purely visual-manual tasks. Less salient effects, however, such as visual effects in an otherwise auditory-manual task, seem to depend more on the fit of their attributes with the current attentional set. Even though we manipulated saliency by contrasting auditory and visual effects, we consider saliency to be a matter of degree and of the particular stimulus-context relations rather than an absolute characteristic of a particular sense modality; but more systematic studies are necessary to elucidate that issue.

Another question is which process exactly is affected by saliency. One possibility is that integration proper depends on some minimal activation of effect codes, which they may reach only if they are either top-down primed because of

[4] Alternatively, participants might have alternated between the integration of auditory and visual effects. However, this should have decreased the response bias for both types of effects, which does not fit the finding that the bias for auditory action effects in the visual-go condition was numerically stronger in Experiment 4 (where alternation might have taken place) than in Experiment 3 (where action effects were all auditory).

their task relevance or particularly salient (see Hommel, 2004). However, a major disadvantage of a selective integration mechanism would be that infants, children, and adults facing a novel task would no longer be able to pick up unpredicted but consistent action-effect relations on the fly—a characteristic of action-effect acquisition that one may consider essential for the development of voluntary action and action skills (Elsner & Hommel, 2001; James, 1890; Piaget, 1946). Another possibility is that action-effect binding is truly automatic and may not even be sensitive to the availability of attentional resources, which would leave binding retrieval as a possible target of our saliency manipulations. Indeed, we cannot exclude that saliency affected the retrieval of just-created action-effect bindings rather than the creation of bindings. That is, stimuli may be more effective to trigger the retrieval of previously created bindings if they are task-relevant or salient. Again, studies on stimulus-response integration suggest that the retrieval of bindings is more sensitive to attentional manipulations than the creation of bindings is (Hommel et al., 2008), which would fit better with a retrieval-based interpretation of saliency effects. Nevertheless, the final word on this matter presupposes a better understanding of how action-effect binding and retrieval processes work, and how they are controlled.

In view of the previous demonstrations of the acquisition of stable action-effect associations on the one side and of the present evidence for transient bindings between actions and effects on the other, it would be tempting to assume that the latter are functional predecessors of the former: The transient coupling of action and effect codes may reflect the presence of reverberatory loops in the sense of Hebb (1949), which again may serve to establish and consolidate more enduring cell assemblies. In other words, binding may represent the first step to long-term memory (cf., Raffone & Wolters, 2001, but see Colzato, Raffone, & Hommel, 2006). However, in the absence of clear-cut evidence that action-effect learning is impossible without binding (and in view of the major methodological challenges demonstrations of such evidence would need to overcome) this is no more than an interesting speculation.

Acknowledgments We are grateful to Wilfried Kunde, who suggested the experimental rationale underlying Experiment 2 and drew our attention to the fact that manipulating action-effect saliency may affect both integration and retrieval of action effects. We are also indebted to two anonymous reviewers whose comments helped improving this article. This research was partially funded by the Max Planck Institute for Psychological Research in Munich, Germany.

References

Beckers, T., De Houwer, J., & Eelen, P. (2002). Automatic integration of non-perceptual action effect features: The case of the associative affective Simon effect. *Psychological Research, 66*, 166–173.

Beringer, J. (1994). ERTS: A flexible software tool for developing and running psychological reaction time experiments on IBM PCs. *Behavior Research Methods. Instruments & Computers, 26*, 368–369.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology, 13*, 90–102.

Bertelson, P. (1963). S-R relationships and reaction times to new versus repeated signals in a serial task. *Journal of Experimental Psychology, 65*, 478–484.

Bogacz, R. (2007). Optimal decision-making theories: Linking neurobiology with behavior. *Trends in Cognitive Sciences, 11*, 118–125.

Colzato, L. S., Fagioli, S., Erasmus, V., & Hommel, B. (2005). Caffeine, but not nicotine enhances visual feature binding. *European Journal of Neuroscience, 21*, 591–595.

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.

Colzato, L. S., van Wouwe, N. C., & Hommel, B. (2007). Feature binding and affect: Emotional modulation of visuo-motor integration. *Neuropsychologia, 45*, 440–446.

Duncan, J. (1996). Cooperating brain systems in selective perception and action. In T. Inui & J. L. McClelland (Eds.), *Attention and performance XVI* (pp. 549–576). Cambridge, MA: MIT Press.

Duncan, J., Humphreys, G., & Ward, R. (1997). Competitive brain activity in visual attention. *Current Opinion in Neurobiology, 7*, 255–261.

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

Elsner, B., & Hommel, B. (2004). Contiguity and contingency in the acquisition of action effects. *Psychological Research, 68*, 138–154.

Elsner, B., Hommel, B., Mentschel, C., Drzezga, A., Prinz, W., Conrad, B., et al. (2002). Linking actions and their perceivable consequences in the human brain. *NeuroImage, 17*, 364–372.

Hebb, D. O. (1949). *The organization of behavior: A neurophysiological theory*. New York: Wiley.

Hoffmann, J., Sebald, A., & Stöcker, C. (2001). Irrelevant response effects improve serial learning in serial reaction time tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 470–482.

Hommel, B. (1993). Inverting the Simon effect by intention: Determinants of direction and extent of effects of irrelevant spatial information. *Psychological Research, 55*, 270–279.

Hommel, B. (1994). Spontaneous decay of response code activation. *Psychological Research, 56*, 261–268.

Hommel, B. (1996). The cognitive representation of action: Automatic integration of perceived action effects. *Psychological Research, 59*, 176–186.

Hommel, B. (1998). Event files: Evidence 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 and Performance, 31*, 1067–1082.

Hommel, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research, 71*, 42–63.

Hommel, B., & Elsner, B. (2009). Acquisition, representation, and control of action. In E. Morsella, J. A. Bargh & P. M. Gollwitzer (Eds.), *Oxford handbook of human action* (pp. 371–398). New York: Oxford University Press.

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

Hommel, B., Memelink, J., Colzato, L. S., & Zmigrod, S. (2008). Attentional control of the creation and retrieval of stimulus-response bindings (submitted).

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*, 849–878.

James, W. (1890). *The principles of psychology*. New York: Dover Publications.

Jolicœur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36*, 138–202.

Lotze, R. H. (1852). *Medicinische Psychologie oder die Physiologie der Seele*. Leipzig: Weidmann'sche Buchhandlung.

Melcher, T., Weidema, M., Eenshuistra, R. M., Hommel, B., & Gruber, O. (2008). The neural substrate of the ideomotor principle: An event-related fMRI analysis. *NeuroImage, 39*, 1274–1288.

Piaget, J. (1946) *La formation du symbole chez l'enfant*. Delachaux & Niestlé.

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.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience, 13*, 766–785.

Reed, P. (1999). Role of a stimulus filling an action-outcome delay in human judgments of causal effectiveness. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 92–102.

Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgment of causality by human subjects. *Quarterly Journal of Experimental Psychology, 41B*, 139–159.

Soetens, E., Boer, L. C., & Hueting, J. E. (1985). Expectancy or automatic facilitation? Separating sequential effects in two-choice reaction time. *Journal of Experimental Psychology: Human Perception & Performance, 11*, 598–616.

Stoet, G., & Hommel, B. (1999). Action planning and the temporal binding of response codes. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1625–1640.

Ziessler, M. (1998). Response-effect learning as a major component of implicit serial learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 962–978.

Ziessler, M., Nattkemper, D., & Frensch, P. A. (2004). The role of anticipation and intention for the learning of effects of self-performed actions. *Psychological Research, 68*, 163–175.

Anticipative control of voluntary action: Towards a computational model

Pascal Haazebroek & Bernhard Hommel

Leiden University
Cognitive Psychology Unit & Leiden Institute for Brain and Cognition
Leiden, The Netherlands

Submitted January 21, 2009

Running Title: Action control

Correspondence:
Bernhard Hommel
Leiden University
Department of Psychology
Cognitive Psychology Unit
Wassenaarseweg 52
2333 AK Leiden, The Netherlands
e-mail: hommel@fsw.leidenuniv.nl

ABSTRACT

Human action is goal-directed and must thus be guided by anticipations of wanted action effects. How anticipatory action control is possible and how it can emerge from experience is the topic of the ideomotor approach to human action. The approach holds that movements are automatically integrated with representations of their sensory effects, so that reactivating the representation of a wanted effect by “thinking of it” leads to a reactivation of the associated movement. We present a broader theoretical framework of human perception and action control—the Theory of Event Coding (TEC)—that is based on the ideomotor principle, and discuss our recent attempts to implement TEC by means of a computational model (HiTEC) to provide an effective control architecture for artificial systems and cognitive robots.

Human behavior is commonly proactive rather than reactive. That is, people do not await particular stimulus events to trigger certain responses but, rather, carry out planned actions to reach particular goals. Planning an action ahead and carrying it out in a goal-directed fashion requires prediction and anticipation: in order to select an action that is suited to reach a particular goal presupposes knowledge about relationships between actions and effects, that is, about which goals can be realized by what action. Under some circumstances this knowledge might be generated ad hoc. For instance, should your behavior ever make a flight attendant to drop you by parachute in a desert, your previously acquired knowledge may be insufficient to select among reasonable action alternatives, so you need to make ad hoc predictions to find out where to turn to. But fortunately, most of the situations we encounter are much more familiar and, thus, much easier to deal with. We often have a rough idea about what actions may be suitable under a given goal and in a particular context, simply because we have experience: we have had and reached the same or similar goals and acted in the same or similar situations before.

How experience with one's own actions generates knowledge that guides the efficient selection of actions, and how humans carry out voluntary actions in general, was the central issue in ideomotor approaches to human action control. Authors like Lotze (1852), Harless (1861), and James (1890) were interested in the general question of how the mere thought of a particular action goal can eventually lead to the execution of movements that reach that goal in the absence of any conscious access to the responsible motor processes (*executive ignorance*). Key to the theoretical conclusion they came up with was the insight that actions are means to generate perceptions (of wanted outcomes) and that these perceptions can be anticipated. If there would be an associative mechanism that integrates motor processes (m) with representations of the sensory effects they produce (e), and if the emerging association between movements and effect representations would be bidirectional (m←→e), reactivating the representation of the effect by voluntarily “thinking of it” may suffice to reactivate the associated motor processes (e→m). In other words, integrating movements and their sensory consequences provides a knowledge base that allows for selecting actions according to their anticipated outcomes—for anticipative action control that is.

After a flowering period in the second half of the 19th century ideomotor approaches were effectively eliminated from the scientific stage (Prinz, 1987; Stock & Stock, 2004). A major reason for that was the interest of ideomotor theoreticians in conscious experience and the relationship between conscious goal representations and unconscious motor behavior, a topic that did not meet scientific criteria in the eyes of the behaviorist movement gaining power in the beginning of the 20th century (cf., Thorndike, 1913). Starting with an early resurrectional attempt by Greenwald (1970), ideomotor ideas have recently regained scientific credibility and explanatory power however. In their Theory of Event Coding (TEC), Hommel, Müsseler, Aschersleben, and Prinz (2001) have even suggested that the ideomotor principle may represent a firm base on which a comprehensive theory of human perception and anticipatory action control can be built. In the following, we will elaborate on what such a theory may look like. In particular, we will briefly discuss the basic principles and basic assumptions of TEC and then go on to describe our recent attempts to implement these principles and assumptions by means of a computational model of human perception and action control—a model we coined HiTEC (Haazebroek & Hommel, submitted).

TEC

The core idea underlying TEC (Hommel et al., 2001) is that perception and action are in some sense the same thing and must therefore be cognitively represented in the same way—the notion of *common coding* (Prinz, 1990). According to the ideomotor principle, action consists in intentionally producing wanted effects, that is, in the execution of motor processes for the sake of creating particular sensory events. In contrast to action, perception is commonly conceived of as the passive registration of sensory input. However, Hommel et al. (2001) argue that this conception is incorrect and misleading, as sensory input is commonly actively produced (Dewey, 1896; Gibson, 1979). For instance, even though visual perception needs light hitting the retina, we actively move our eyes, head, and body to make sure that our retina is hit by the light that is reflecting the most interesting and informative events. That is, we actively search for the information we are interested in and move our receptive surfaces to optimize the intake of that information. This is even more obvious for the tactile sense, as almost nothing would be perceived by touch without systematically moving the sensor surface across the objects of interest. Hence, we perceive by executing motor processes for the sake of creating particular sensory events. Obviously, this is exactly the way we just defined action, which implies that action and perception are one process.

The second central assumption of TEC is that cognitive representations are *composites* of feature codes (Hommel, 2004). Our brain does not represent events through individual codes or neurons but by widely distributed feature networks. For instance, the visual cortex consists of numerous representational maps coding for various visual features, such as color, orientation, shape, or motion (DeYoe & Van Essen, 1988) and similar feature maps have been reported for other modalities. Likewise, action plans are composites of neural networks coding for various action features, such as the direction, force, or distance of manual actions (Hommel & Elsner, 2009). One implication of the assumption that cognitive event representations are composites is that binding operations are necessary to integrate the codes referring to the same event, and another is that different events can be related to, compared with, or confused with each other based on the features they do or do not share. For instance, TEC implies that stimuli and responses can be similar to each other, in the sense that the binding representing the stimulus and the binding representing the response can include the same features, such as location or speed, and can thus prime each other (which for instance explains effects of stimulus-response compatibility) or interact in other ways.

The third main assumption of TEC is that the cognitive representations that underlie perception and action planning code for *distal* but not *proximal* aspects of the represented events (Prinz, 1992). In a nutshell, this means that perceived and produced events are coded in terms of the features of the external event *as external event* (i.e., as objectively or inter-subjectively definable) but not with respect to the specifics of the internal processing, such as retinal or cortical coding characteristics, or particular muscle parameters. This terminology goes back to Heider (1926, 1930), who discussed the problem that our conscious experience refers to objective features of visual objects (the distal attributes), even though the intermediate processing steps of the physical image on the retina and the physiological response to it (the proximal attributes) are not fully determined by the distal attributes. Brunswik (1944) extended this logic to action and pointed out that goal representations refer to distal aspect of the goal event and, thus, do not fully determine the proximal means to achieve it.

To summarize, TEC assumes that perceived events are represented by activating and integrating feature codes—codes that represent the distal features of the event. Given that perceptions are actively produced, these bindings are likely to also include action features, that is, codes that represent the features of the action used to produce that perception. In turn, action plans are integrated bindings of codes representing the distal features of the action. As actions are carried out to create sensory events, action plans also comprise of feature codes referring to these events. In other words, both perceived and produced events are represented by sensorimotor bindings or “event files” (Hommel, 2004). However, not all features of a perceived or a produced event are relevant in a particular context. To account for that, TEC assumes that feature codes are “intentionally weighted” according to the goal or task at hand. For instance, if you are searching for a particular color, or if what matters for your actions is the location of your fingertip, color and location codes would be weighted higher, respectively, and thus affect perception and action planning more strongly. TEC was very helpful in interpreting and integrating available findings in a coherent manner, as well as in stimulating numerous experiments and studies on various topics and perception-action phenomena. However, as Hommel et al. (2001) pointed out, TEC only provides a general framework and the theoretical concepts needed to get a better understanding of perception, action, and their relationship. Deeper insight and theoretical advancement calls for more detail and additional assumptions. To meet this challenge we began developing HiTEC, a computational implementation of TEC’s basic principles and assumptions. In the following, we provide a brief overview of the main strategies guiding our implementation, but refer to Haazebroek and Hommel (submitted) for a broader treatment.

HiTEC

HiTEC (Haazebroek & Hommel, submitted) is an attempt to translate the theoretical framework of TEC (Hommel et al, 2001) into a runnable computational model. Our ambition is to develop a broad, cognitive architecture that can account for a variety of empirical effects related to stimulus-response translation and that can serve as a starting point for a novel control architecture for cognitive robots in the PACO-PLUS project (www.paco-plus.org).

From a modeling perspective TEC provides a number of constraints; some of them enforce structural elements while others impose the existence of certain processes. First, we describe the general structure of HiTEC. Next, we elaborate on the processes operating on this structure, following the two-stage model (Elsner and Hommel, 2001) for the acquisition of voluntary action control. Finally, we discuss how the mechanisms of HiTEC might operate in a real life scenario and show that anticipation plays a crucial role in quickly generating and controlling appropriate responses.

===== FIGURE 1 ABOUT HERE =====

HiTEC’S STRUCTURE AND REPRESENTATIONS

HiTEC is architected as a connectionist network model that uses the basic building blocks of parallel distributed processing (PDP; e.g., McClelland, 1992; Rumelhart, Hinton, & McClelland, 1986). In a PDP network model processing occurs through the interactions of a large number of interconnected elements called units or nodes. Nodes may be organized into higher structures, called modules, each containing a number of nodes. Modules may be part of a larger processing pathway. Pathways may interact in the sense that they can share common modules.

Each node has an activation value indicating local activity. Processing occurs by propagating activity through the network; that is, by propagating activation from one node to the other, via weighted connections. When a connection between two nodes is positively weighted, the connection is excitatory and the nodes will increase each other’s activation. When the connection is negatively weighted, it is inhibitory and the nodes will reduce each other’s activation. Processing starts when one or more nodes receive some sort of external input. Gradually, node activations will rise and propagate through the network while interactions between nodes control the flow of processing. Some nodes are designated output nodes. When activations of these nodes reach a certain threshold (or when the time allowed for processing has passed), the network is said to produce the corresponding output(s).

In HiTEC, the elementary units are codes. As illustrated in Figure 1, codes are organized into three main systems: the sensory system, the motor system and the common coding system. Each system will now be discussed in more detail.

Sensory System

As already mentioned, the primate brain encodes perceived objects in a distributed fashion: different features are processed and represented across different cortical maps (e.g., Cowey, 1985; DeYoe & Van Essen, 1988). In HiTEC, different modalities (e.g., visual, auditory) and different dimensions within each modality (e.g., visual color and shape, auditory location and pitch) are processed and represented in different sensory maps. Each sensory map is a module containing a number of sensory codes that are responsive to specific sensory features (e.g., a specific color or a specific pitch). Note that Figure 1, shows only two sensory codes per map for clarity.

In the visual brain, there are two major parallel pathways (Milner & Goodale, 1995) that follow a common preliminary basic feature analysis step. The ventral pathway is seen as crucial for object recognition and consists of a hierarchy of sensory maps coding for increasingly complex features (from short line segments in the lower maps to complex shapes in higher maps) and increasingly large receptive field (from a small part of the retina in the lower maps to anywhere on the retina in higher maps). The second pathway, the dorsal pathway, is seen as crucial for action guidance as it loses color and shape information but retains information about contrast, location of objects, and other action-related features.

In HiTEC, a common visual sensory map codes for basic visual parts of perceptual events. This common basic map projects to both the ventral and the dorsal pathways. The ventral pathway consists of sensory maps coding for combinations (such as more specific shapes) or abstractions (e.g., object color). The dorsal pathway is currently simply a sensory map coding for visual location—to be extended for processing other action-related features in a later version of HiTEC.

Distributed processing allows a system to dramatically increase its representational capacity as it no longer requires each combination of features to have its own dedicated representational structure but can rather encode a specific combination on demand in terms of activating a collection of constituting feature structures. On the downside, in typical scenarios, this inevitably results in binding problems (Treisman, 1996). For instance, when multiple objects are perceived and they are both represented in terms of activating the structures coding for their constituting features, how to tell which feature belongs to which object? This clearly calls for an integration mechanism that can tell them apart.

Recent studies in the visual modality have shown that this problem can, partly, be solved by employing local interactions between feed-forward and feed-back processes in the ventral and dorsal pathways (Van der Velde & De Kamps, 2001). It is true that higher ventral sensory maps do not contain information on location and that higher dorsal sensory maps do not contain information on object shape or color, but these pathways can interact using the common basic visual feature map as a visual blackboard (Van der Velde, De Kamps, & Van der Voort van der Kleij, 2004). For instance: when a specific color is activated in a higher sensory map, it can feed back activation to lower sensory maps, thereby modulating the activity of these sensory codes in a way that those codes that code for simple parts of this color are enhanced. This can modulate the processing in the dorsal pathway as well resulting in enhanced activation of those codes in the location map that code for the location(s) of objects of the specified color.

This principle also works the other way round: activating a specific location code in the location map can modulate the sensory codes in the lower sensory maps that code for simple parts at this location. This can modulate the processing in the ventral pathway, resulting in enhanced activation of the more complex or abstract features of the object at the specified location. In HiTEC, this is the way the visual sensory system can be made to enhance the processing of objects with specific features or on a specific location. For now, we assume the following sensory maps in the HiTEC architecture: visual basic features map, visual color map, visual shape map, visual location map, auditory pitch map, auditory location map, tactile effector (i.e., hands or feet) map and tactile location map.

Motor System

The motor system contains motor codes, referring to proximal aspects of movements. Motor codes can also be organized in maps, following empirical evidence that suggests distributed representations at different cortical locations in the motor domain (e.g., Andersen, 1988; Colby 1998). For example, cortical maps can be related to effector (e.g., eye, hand, arm, foot) or movement type (e.g., grasping, pointing). It makes sense to assume that there is some sort of hierarchical structure as well in motor coding. However, in the present version of HiTEC, we consider only one basic motor map with a set of motor codes. As our modeling efforts in HiTEC evolve, its motor system may be extended further.

It is clear that motor codes, even when structured in multiple maps, can only specify a rough outline of the motor action to be performed as some parameters depend strongly on the environment. For instance, when grasping an object, the actual object location is not represented by a motor code (this would lead to an explosion of the number of necessary motor codes, even for a very limited set of actions). So it makes sense to interpret a motor program as a blueprint of a motor action that needs to be filled in with this specific, on line, information, much like the schemas put forward by Schmidt (1975) and Glover (2004). In our discussion of HiTEC processes we will discuss this issue in more detail.

Common Coding System

According to TEC both perceived events and action generated events are coded in one common representational domain (Hommel et al, 2001). In HiTEC, this domain is the common coding system that contains common feature codes. Feature codes refer to distal features of objects, people and events in the environment. Example features are distance, size and location, but on a distal, descriptive level, as opposed to the proximal features as coded by the sensory codes and motor codes.

Feature codes may be associated to both sensory codes and motor codes and are therefore truly sensorimotor. They can combine information from different modalities and are in principle unlimited in number. Feature codes are not given but they evolve and change. In HiTEC simulations, however, we usually assume a set of feature codes to be present initially, to bootstrap the process of extracting sensorimotor regularities in interactions with the environment.

Feature codes are contained in feature dimensions. As feature dimensions may be enhanced as a whole, for each dimension an additional dimension code is added that is associated with each feature code within this dimension. Activating this code will spread activation towards all feature codes within this dimension, making them more sensitive to stimulation originating from sensory codes.

Associations

In HiTEC, codes can become associated, both for short term and for long term. Short term associations between feature codes reflect that these codes 'belong together in the current task or context’ and their binding is actively maintained in working memory. In Figure 1, these temporary bindings are depicted as dashed lines. Long term associations can be interpreted as learned connections reflecting prior experience. For now, we do not differentiate between episodic and semantic memory—even though later versions are planned to distinguish between a “literal” episodic memory that stores event files (see below) and a semantic memory that stores rules abstracted from episodic memory (O'Reilly & Norman, 2002). At present, both types of experience are modeled as long term associations between (any kind of) codes and are depicted as solid lines in Figure 1.

Event file

Another central concept in the theory of event coding is the event file (Hommel, 2004). In HiTEC, the event file is modeled as a structure that temporarily associates to feature codes that 'belong together in the current context’ in working memory. The event file serves both the perception of a stimulus as well as the planning of an action. Event files can compete with other event files.

HiTEC’S PROCESSES

How do associations between codes come to be? What mechanisms result of their interactions? And how do these mechanisms give rise to anticipation based, voluntary action control? Elsner and Hommel (2001) proposed a two-stage model for the acquisition of voluntary action control. At the first stage, the cognitive system observes and learns regularities in motor actions and their effects. At the second stage, the system uses the acquired knowledge of these regularities to select and control its actions. For both stages, we now discuss in detail how processes take place in the HiTEC architecture. Next, we discuss some additional process related aspects of the architecture.

Stage 1: acquiring action-effect associations

The framework of event coding assumes that feature codes are grounded representations as they are derived by abstracting regularities in activations of sensory codes. However, the associations between feature codes and motor codes actually signify a slightly different relation: feature codes encode the (distal) perceptual effect of the action that is executed by activating the motor codes. Following the ideomotor principle, the cognitive system has no innate knowledge of the actual motor action following the activation of a certain motor code. Rather, motor codes need to become associated with their perceptual action effects so that by anticipating these effects, activation can propagate via these associations to those motor codes that actually execute the corresponding movement.

Infants typically start off with a behavioral repertoire based on stimulus-response (SR) reflexes (Piaget, 1952). As the infant exhibits these stimulus-response reflexes, as well as random behaviors (e.g., motor babbling), its cognitive system learns the accompanying response-perceptual effect (RE) regularities that will serve as some sort of database of ‘what action achieves what environmental effect’. Following Hommel (1996), we assume that any perceivable action effect is automatically coded and integrated into an action concept, which is, in the HiTEC architecture, an event file consisting of feature codes. Although all effects of an action become integrated automatically, intentional processes do affect the relative weighting of integrated action effects—TEC's intentional-weighting principle.

Taken together, action – effect acquisition is modeled in HiTEC as follows: motor codes m_i are activated, either because of some already existing associations or simply because of network noise. This leads to a change in the environment (e.g., the left hand suddenly touches a cup) which is picked up by sensory codes s_i. Activation propagates from sensory codes towards feature codes f_i. And eventually, these feature codes are integrated into an event file e_i which acts as an action concept. Subsequently, the cognitive system learns associations between the feature codes f_i belonging to this action concept and the motor code m_i that just led to the executed motor action. Crucially, task context can influence the learning of action effects. Not by selecting which effects are associated but by weighting the different effect features. Nonetheless, this is an interactive process that does not exclude unintended but utterly salient action effects to become involved in strong associations as well.

Stage 2: using action effect associations

Once associations between motor codes and feature codes exist, they can be used to select and plan voluntary actions. Thus, by anticipating desired action effects, feature codes become active. Now, by integrating the feature codes into an action concept, the system can treat the features as constituting a desired state and propagate their activation towards associated motor codes. Crucially, anticipating certain features needs integration to tell them apart from the features that code for the currently observed environment. Once integrated, the system has 'a lock' on these features and can use these features to select the right motor action.

Initially, multiple motor codes m_i may become active as they typically fan out associations to multiple feature codes f_i. However, some motor codes will have more associated features that are also part of the active action concept and some of the m_i - f_i

associations may be stronger than others. Taken together, the network will – in PDP fashion – converge towards one strongly activated motor code m_i which will lead to the selection of that motor action.

In addition to the mere selection of a motor action, feature codes also form the actual action plan that specifies (in distal terms) how the action should be executed: namely, in such a way the intended action effects are realized. By using anticipated action effect to choose an action, the action actually is selected because the cognitive system intended this, not because of a reflex to some external stimulus. Thus, in HiTEC, using anticipation is the key to voluntary action.

Task context

Task context can modulate both action-effect learning and the usage of these links. This can help focus processing to action alternatives that 'make sense' in the current context. In real life this is necessary as the action alternatives are often rather unconstrained. Task context comes in different forms. One is the overall environment, the scene context in which the interaction takes place. The cognitive system may just have seen other objects in the room, or the room itself, and feature codes that code for aspects of this context may still have some activation. This can, in principle, influence action selection. As episodic and semantic memory links exist as well, this influence may also be less salient: the presence of a certain object might recall memories of previous encounters or similar contexts that influence action selection in the current task.

A task can also be very specific, as given by a tutor or instructor in terms of a verbal description. In HiTEC, it is assumed that feature codes can be activated by means of verbal labels. Thus, when a verbal task is given, this could directly activate feature codes. The cognitive system integrates these codes into an event file that is actively maintained in working memory. For example, when approached with several options to respond differently to, different event files e_i are created for the different options. Due to the mutual inhibitory links between event files, they will compete with each other. Because of the efficiency the cognitive system can now display, one could state that a cognitive reflex has been prepared (Hommel, 2000) that anticipates certain stimuli features. The moment these features are actually perceived, the reflex ‘fires’ and - by propagating activation to event codes and subsequently to other feature codes - quickly anticipates the correct action effects, which results in the selection and execution of the correct motor action.

Online vs offline processing

In HiTEC, action selection and action planning are interwoven, but on a distal feature level. This leaves out the necessity of coding every minute detail of the action, but restricts action planning to a ballpark idea of the movement. Still, a lot has to be filled in by on line information. Currently, this falls outside the scope of HiTEC, but one could imagine that by activating distal features, the proximal sensory codes can be top down moderated to ‘focus their attention’ towards specific aspects of the environment (e.g., visual object location), see Hommel (in press). In addition, actions need still not to be completely specified in advance, as they are monitored and adjusted while they are performed—which in humans seems to be the major purpose of dorsal pathways (Milner & Goodale, 1995)

Action monitoring

The anticipated action effects are a trigger for action selection, but also form an expectation of the perceptual outcome of the action. Differences between this expectation and reality lead to adjusting the action on a lower sensorimotor level than is currently modeled in HiTEC. What matters now, is that the feature codes are interacting with the sensory codes, making sure that the generated perception is within the set parameters, as determined by the expected action outcome. If this is not (well enough) the case, the action should be adjusted.

However, when a discrepancy of this expectation drastically exceeds 'adjustment thresholds’, it may actually trigger action effect learning (stage 1). Apparently, the action-effect associations were unable to deliver an apt expectation of the actual outcome. Thus, anticipating the desired outcome falsely led to the execution of this action. This may trigger the system to modify these associations, so that the motor codes become associated with the correct action effect features.

Crucially, having anticipations serve as expectations, the system is not forced into two distinct operating modes (learning vs. testing). With anticipation as retrieval cue for action selection and as expectation of the action outcome, the system has the means to self-regulate its learning by making use of the discrepancy between actual effects and these anticipations.

===== FIGURE 2 ABOUT HERE =====

EXAMPLARY SCENARIO: RESPONDING TO TRAFFIC LIGHTS

In order to clarify the co-operation of the different processes and mechanisms in HiTEC, the following example real life scenario is presented: learning to respond to traffic lights. In this example, s_i denotes sensory codes, f_i denotes feature codes and m_i denotes motor codes in the HiTEC architecture. Figure 2 shows a scenario-specific version of the HiTEC architecture.

Action effect acquisition

Let's say you are a student driver who has never paid attention to the front seat before and this is your first driving lesson. You climb behind the steering wheel and place your feet above the pedals. Now, the instructor starts the car for you and you get the chance of playing around with the pedals. After a while, you get the hang of it: it seems that pressing the right pedal results in a forward movement of the car, and pressing the left one puts the car on hold.

From a HiTEC perspective, you just have tried some motor codes and learned that m_1 (pressing the gas pedal) results in a forward motion, coded by $f_{forward}$ and m_2 in standing still, coded by f_{stop}. In other words: you acquired these particular action-effect associations. Note that we assume that you have been able to walk before, so it is fair to say that $f_{forward}$ and f_{stop} are already present as feature codes in your common coding system.

Using action effect associations

Now, in your next lesson you actually need to take cross roads. The instructor tells you to pay attention to these colored lights next to the road. When the red light is on, you should stop, and when the green light is on, you can go forward.

In HiTEC, this verbal instruction is modeled as creating two event files that hold short term associations in working memory: $e_{stop\ for\ red\ light}$ for the 'stop' condition, and $e_{go\ at\ green\ light}$ for the 'forward' condition. The event file $e_{stop\ for\ red\ light}$ contains bindings of feature codes f_{red}, $f_{traffic\ light}$, f_{stop} and the event file $e_{go\ at\ green\ light}$ relates to the feature codes f_{green}, $f_{traffic\ light}$, $f_{forward}$.

These event files are activated and their activation spreads to their associated feature codes which will become increasingly receptive for interaction with related sensory codes. In addition to the specific features, the feature dimensions these features are contained in (d_{color}, d_{motion}) are weighted as well. The anticipation of traffic lights also serves as a retrieval cue for prior experience with looking at traffic lights. As traffic lights typically stand at the side of the road, one could expect associations between $f_{traffic\ light}$ and $f_{side\ of\ road}$ to exist in episodic or semantic memory. Consequently, anticipating a traffic light activates $f_{traffic\ light}$ and propagates activation automatically towards $f_{side\ of\ road}$, which makes the system more sensitive to objects located on the side of the road.

Ok, there it goes... you start to drive around, take some turns, and there it is… your very first cross road with traffic lights!

Now, from a HiTEC perspective, the following takes place: the visual scene consists of a plethora of objects, like road signs, other cars, houses and scenery, and of a cross road with traffic lights at the side. The sensory system encodes the registration of these objects by activating the codes in the sensory maps. This leads to the classical binding problem: multiple shapes are registered, multiple colors and multiple locations. However, we now have a top down 'special interest' for traffic lights. As mentioned above, this has resulted in increased sensitivity of the $f_{traffic\ light}$ feature code, that now receives some external stimulation from related sensory codes. Also, from prior experience we look more closely at $f_{side\ of\ road}$ locations in the sensory location maps.

The interaction between this top down sensitivity and the bottom up external stimulation results in an interactive process where the sensory system uses feedback signals to the lower level visual maps where local interactions result in higher activation of those sensory codes that code for properties of the traffic light, including its color. In the visual map for object color, the traffic light color will be more enhanced than colors relating other objects. On the feature code level, the color dimension already was enhanced because of the anticipation of features in the d_{color} dimension, resulting in fast detection of f_{red} or f_{green}.

Meanwhile, the event files $e_{stop\ for\ red\ light}$ and $e_{stop\ for\ red\ light}$ are still in competition. When the sensory system collects the evidence, activation propagates towards feature codes and event codes, quickly converging into a state that where either $f_{forward}$ or f_{stop} is activated more strongly than the other. This activation is propagated towards the motor codes m_1 or m_2 via associations learned in your first drivers lesson. This results in the selection and execution of the correct motor action.

It is clear that by preparing the cognitive system for perceiving a traffic light color and producing a stop-or-go action allows the system to effectively attend its resources to the crucial sensory input and already pre-anticipate the possible action outcome. This way, upon perceiving the actual traffic light color, the system can quickly respond with the correct motor action.

Luckily, for your safety and that of all your fellow drivers on the road, practicing this task long enough will also result in long term memory bindings between f_{red}, $f_{traffic\ light}$ and f_{stop} that will also be retrieved during action selection and bias you towards pressing the brake pedal, even when no instructor is sitting next to you.

CONCLUSIONS

We have introduced HiTEC’s three main modules: the sensory system, the motor system, and the emergent common coding system. These systems interact with each other. In the common coding system anticipations are formed that have a variety of uses in the architecture, allowing the system to be more flexible and adaptive. In action selection, anticipation acts as a rich retrieval cue for associated motor programs. At the same time, forming this anticipation reflects the specification of an action plan that can be used during action execution.

One of the drawbacks of creating anticipations is that it might not be worth the costs (Butz & Pezzulo, 2008). However, from a real life scenario perspective, the number of possible action alternatives is enormous. Creating anticipations at a distal level seems as a necessity to constrain the system in its actions to select from. Doing this, as we propose in HiTEC, not only aids action selection but also delivers the rudimentary action plan at the same time.

Another concern often mentioned is the inaccuracy of predictions. Following the framework of event coding, events – including action plans – are coded in distal terms that abstract away from the proximal sensory values. Only inaccuracies on the distal level could disturb the use of anticipations in action selection and planning. The feature codes on this distal level are based on sensorimotor regularities that are stable over time. Thus minor inaccuracies in sensors should be relatively easily overcome.

Actions are usually selected and planned in a task context. When forced with different behavioral alternatives to choose from, multiple anticipations of features are created and compete with each other. When features are actually perceived, anticipatory activation quickly propagates to the correct action effects, which results in the selection and execution of the correct motor action.

In action monitoring, anticipation serves as the representation of expected and desired action effects that helps adjusting the movement during action execution. In action evaluation, this expectation acts as a set of criteria for success of the action. If the actual action effect can no longer – on a lower sensorimotor level - be adjusted to fulfill the expected action effect, the existing action-effect associations are considered insufficient and learning is triggered. During action-effect learning, anticipation also may weight the different action effect features in the automatic integration into action concepts, influencing the action-effect association weights.

In conclusion, anticipation plays a crucial role in virtually all aspects of action control within the HiTEC architecture. Just as it does in real life.

REFERENCES

Andersen, R. A. (1988). The neurobiological basis of spatial cognition: Role of the parietal lobe. In: *Spatial cognition: Brain bases and development,* ed. J. Stiles-Davis, M. Krtichivsky & U. Belugi. Erlbaum.

Brunswik, E. (1944). Distal focussing of perception: Size constancy in a representative sample of situations. *Psychological Monographs, 56, No. 1.*

Butz, M. V. & Pezzulo, G. (2008). Benefits of Anticipations in Cognitive Agents. In G. Pezzulo et al. (Eds.): *The Challenge of Anticipation*, pp. 45-62.

Colby, C. L. (1998). Action-oriented spatial reference frames in the cortex. *Neuron 20,* 15-20.

Cowey, A. (1985). Aspects of cortical organization related to selective attention and selective impairments of visual perception: A tutorial review. In M. I. Poster & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 41-62). Hillsdale, NJ: Erlbaum.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological Review, 3*, 357-370.

DeYoe, E. A., & Van Essen, D. C. (1988). Concurrent processing streams in monkey visual cortex. *Trends in Neuroscience, 11,* 219-226.

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

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, Houghton Mifflin.

Glover, S. (2004). Separate visual representations in the planning and control of action. *Behavioral and Brain Sciences, 27,* 3-24.

Greenwald, A. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideomotor mechanism. *Psychological Review, 77,* 73-99.

Harless, E. (1861). Der Apparat des Willens. *Zeitschrift fuer Philosophie und philosophische Kritik, 38*, 50-73.

Haazebroek, P., & Hommel, B. (submitted). HiTEC: A computational model of the interaction between perception and action.

Heider, F. (1926/1959). Thing and medium. *Psychological Issues, 1959, Monograph 3* (original work published 1926).

Heider, F. (1930/1959). The function of the perceptual system. *Psychological Issues*, *1959, Monograph*, 371-394. (original work published 1930).

Hommel, B. (1996). The cognitive representation of action: Automatic integration of perceived action effects. *Psychological Research, 59*, 176-186.

Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus-response translation. In S. Monsell & J. Driver (eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 247-273). Cambridge, MA: MIT Press.

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

Hommel, B. (in press). Grounding attention in action control: The intentional control of selection. In B.J. Bruya (ed.), *Effortless attention: A new perspective in the cognitive science of attention and action*. Cambridge, MA: MIT Press.

Hommel, B., & Elsner, B. (2009). Acquisition, representation, and control of action. In E. Morsella, J. A. Bargh, & P. M. Gollwitzer (eds.), *Oxford handbook of human action* (pp. 371-398). New York: Oxford University Press.

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

James, W. (1890). *The principles of psychology.* New York: Dover Publications.

Lotze, R. H. (1852). *Medicinische Psychologie oder die Physiologie der Seele*. Leipzig: Weidmann'sche Buchhandlung.

McClelland, J. L. (1992). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence and cognitive neuroscience – A Silver Jubilee Volume*. Cambridge, MA: MIT Press.

Milner, A. D. & Goodale, M. A. (1995). *The visual brain in action*. Oxford University Press.

O'Reilly, R.C. & Norman, K.A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences, 6,* 505-510

Piaget, J. 1952. The origins of intelligence in childhood. International Universities Press.

Prinz, W. (1987). Ideo-motor action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action*. Hillsdale, NJ: Erlbaum.

Prinz, W. (1990). A common coding approach to perception and action. In O. Neumann, & W. Prinz (Eds.), *Relationships between perception and action* (pp. 167-201). Berlin: Springer Verlag.

Prinz, W. (1992). Why don't we perceive our brain states? *European Journal of Cognitive Psychology, 4,* 1-20.

Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 45-76). Cambridge, MA: MIT Press.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning.

Psychological Review, 82, 225-260.

Stock, A. & Stock, C. (2004). A short history of ideo-motor action. *Psychological Research, 68,* 176-188.

Thorndike, E. L. (1913). Ideo-motor action. *Psychological Review, 20,* 91-106.

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

Van der Velde, F., & De Kamps, M. (2001). From knowing what to knowing where: Modeling object-based attention with feedback disinhibition of activation. *Journal of Cognitive Neuroscience, 13 (4),* 479–491.

Van der Velde, F., De Kamps,M., & Van der Voort van der Kleij, G. (2004). Clam: Closed-loop attention model for visual search. *Neurocomputing, 58-60,* 607–612.

ACKNOWLEDGMENTS

Support for this research by the European Commission (PACO-PLUS, IST-FP6-IP-027657) is gratefully acknowledged.

FIGURE

Figure 1. General architecture of HiTEC

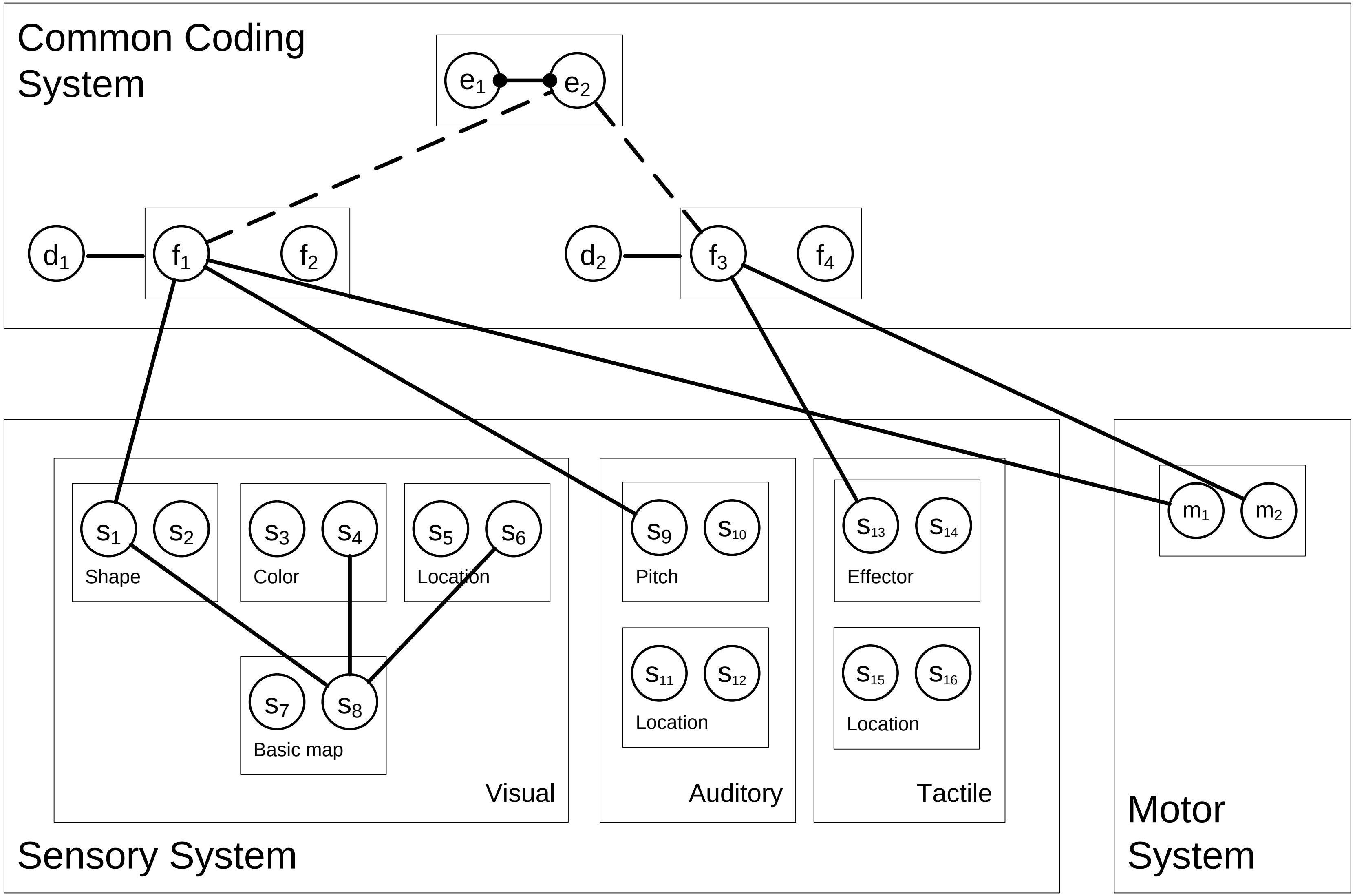
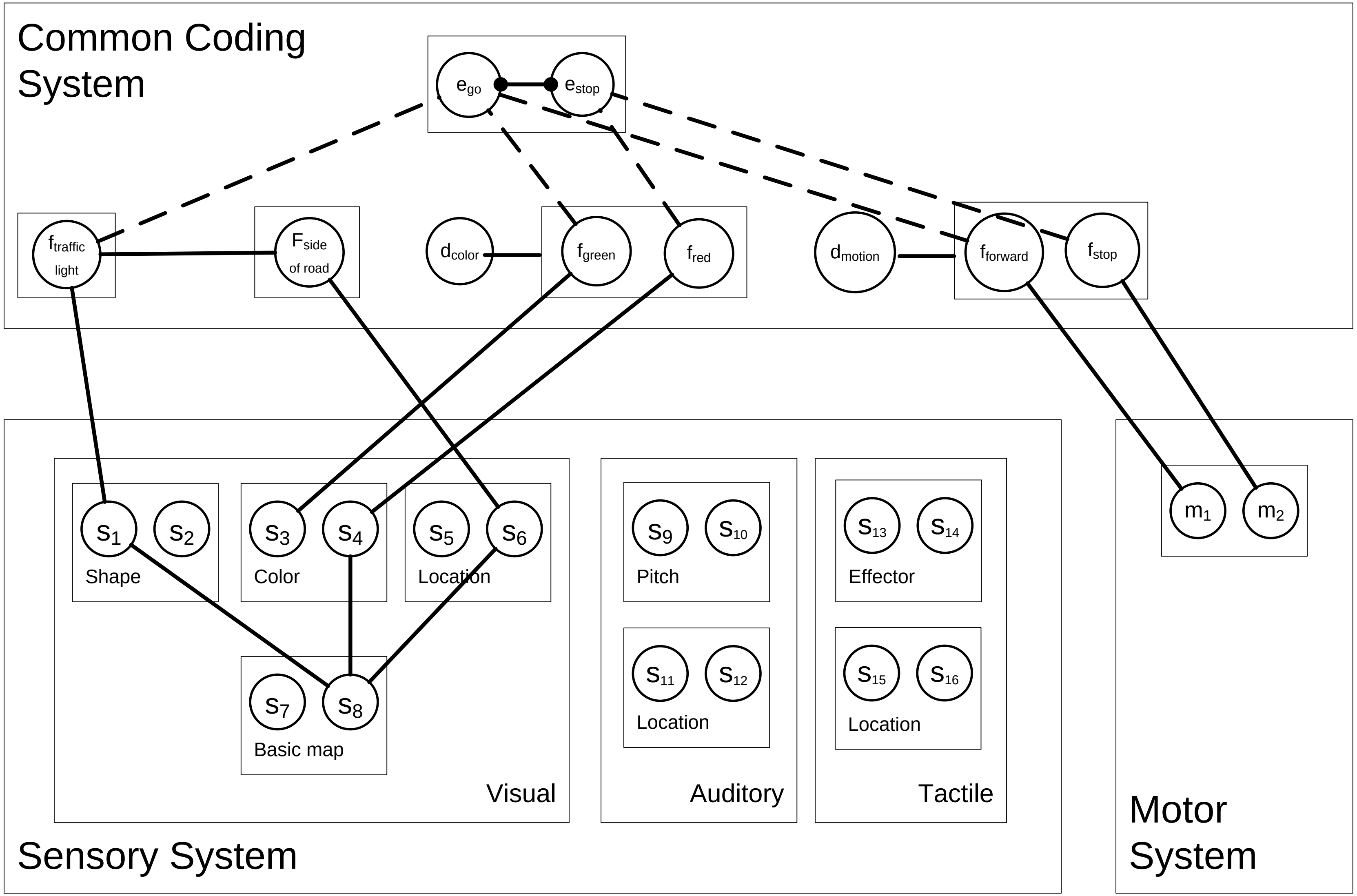


Figure 2. Learning to respond to traffic lights in HiTEC



Psychological Research
DOI 10.1007/s00426-008-0163-5

ORIGINAL ARTICLE

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

Sharon Zmigrod · Michiel Spapé · Bernhard Hommel

Received: 19 May 2008 / Accepted: 8 September 2008

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

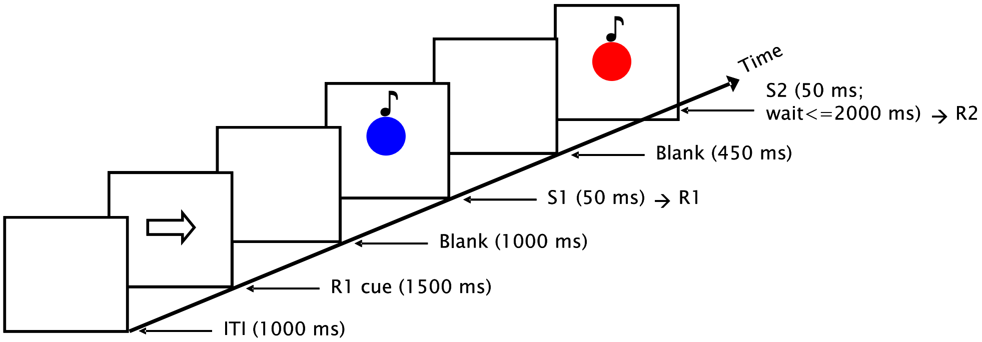


Fig. 1 Sequence of events in Experiment 1. A visual response cue signaled a left or right mouse button click (R1) that was to be delayed until presentation of an audiovisual stimulus S1 (S1 is used as a detection signal for R1). The audiovisual stimulus S2 appeared 450 ms after R1. S2 signaled R2, a 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

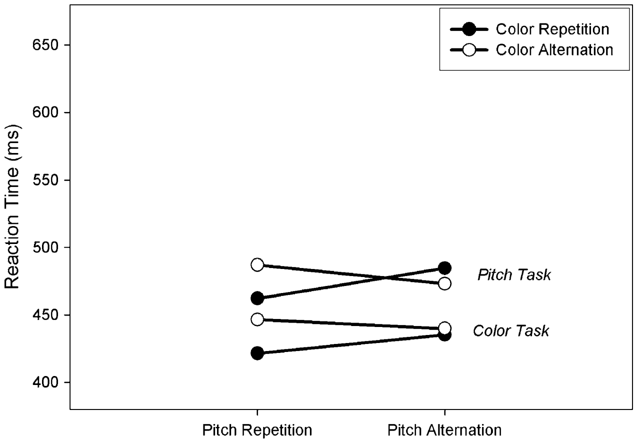


Fig. 2 Reaction times of R2 in Experiment 1 as a function of repetition versus alternation of the stimuli (S1–S2) of visual feature color and auditory feature pitch, regardless of the response

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

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

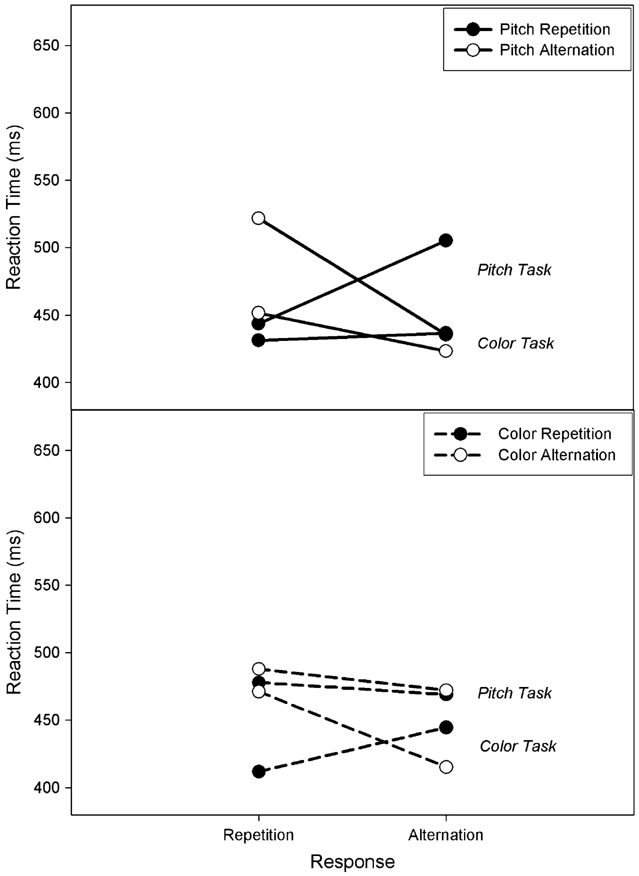


Fig. 3 Reaction times of R2 in Experiment 1 for repetition versus alternation of the stimuli in the auditory feature pitch and the visual feature color, as a function of response repetition (vs. alternation) and task

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

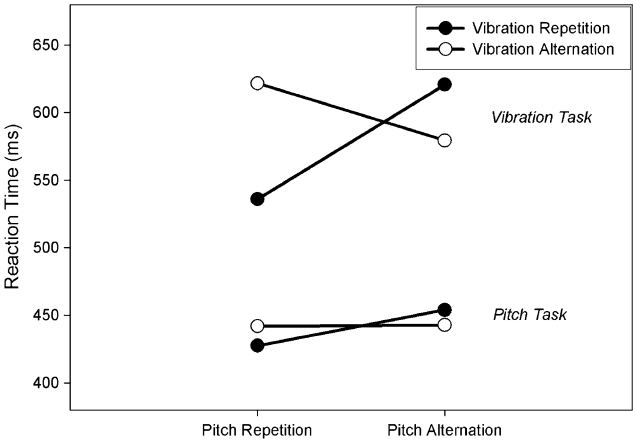


Fig. 4 Reaction times of R2 in Experiment 2 as a function of repetition versus alternation of the stimuli (S1–S2) of tactile feature vibration and auditory feature pitch, and task

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

*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$

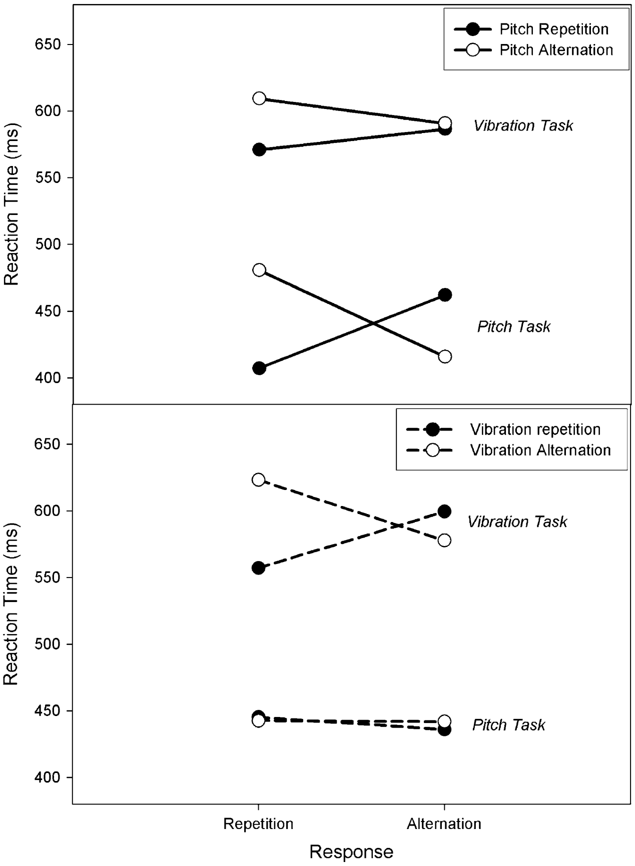


Fig. 5 Reaction times of R2 in Experiment 2 for repetition versus alternation of the stimuli in the auditory feature pitch and the tactile feature vibration, as a function of response repetition (vs. alternation) and task

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*.