Journal of Integrative Neuroscience, Vol. 4, No. 2 (2005) 1–18
© Imperial College Press
ICP Imperial College Press
www.icpress.co.uk

Research Report

EVENT BASED SELF-SUPERVISED TEMPORAL INTEGRATION FOR MULTIMODAL SENSOR DATA

EMILIA I. BARAKOVA*

RIKEN Brain Science Institute,
2-1 Horosawa,Wako-shi, Saitama,351-0198, Japan
emilia@brain.riken.jp

TINO LOURENS

Honda Research Institute Japan Co., Ltd.,
8-1 Honcho, Wako-shi, Saitama, 351-0114, Japan
tino@jp.honda-ri.com

Received 20 October 2004
Revised 18 April 2005

A method for synergistic integration of multimodal sensor data is proposed in this paper. This method is based on two aspects of the integration process: (1) achieving synergistic integration of two or more sensory modalities, and (2) fusing the various information streams at particular moments during processing. Inspired by psychophysical experiments, we propose a self-supervised learning method for achieving synergy with combined representations. Evidence from temporal registration and binding experiments indicates that different cues are processed individually at specific time intervals. Therefore, an event-based temporal co-occurrence principle is proposed for the integration process. This integration method was applied to a mobile robot exploring unfamiliar environments. Simulations showed that integration enhanced route recognition with many perceptual similarities; moreover, they indicate that a perceptual hierarchy of knowledge about instant movement contributes significantly to short-term navigation, but that visual perceptions have bigger impact over longer intervals.

Keywords: Multimodal integration; robotics; navigation; proprioception.

1. Introduction

The world around us supplies huge amounts of information continuously from which living organisms extract the knowledge and awareness we need for survival. A fundamental cognitive feature that makes this possible is the brain's ability to integrate all the various sensory inputs into a coherent representation of its environment. By

*Corresponding author.

analogy, robots are designed to continuously record large amounts of data using various sensors, but their effectiveness suffers from a major design flaw. They lack reference of how the information from the different sensory streams can be integrated into consistent representations.

Multimodal integration for navigation has been studied from many different perspectives. For instance, some studies have taken a purely computational approach for their multimodal systems [19, 30]. Other studies took inspiration from observable animal or human behaviors [3, 39, 24, 28, 41]. A third approach models neural mechanisms in areas of the brain associated with multimodal integration [1, 2, 3, 11, 30, 31, 35, 38].

Our novel solution for this multimodal sensory integration problem for robot navigation task was inspired by psychophysical experiments connecting visual and idiothetic information. While navigating around an area, idiothetic information is internally generated as the body moves through space [9, 25]. This information can be derived from proprioceptive sensory streams about own movements and motor efferent signals. Vestibular information, which follows the change in linear or rotational movement velocity, is another source of idiothetic information. We integrated view-based and velocity sensory information in our robotic implementations.

In general, two types of problems have to be solved for multimodal integration: how and when should sensory cues be fused and how to obtain a synergistic multimodal integration. Synergistic integration should produce more information from the integrated representation than is evident from information generated in the separate modalities.

The first set of problems concerns the representational and technical aspects of the actual fusion process. The method we proposed represents continuous sensory information dynamically, by encoding the temporal history of sensor readings. This encoding is a simple model of short-term memory. The method assumes that different percepts unify in the brain, as suggested from temporal registration and binding experiments showing that the information from one type of sensors is processed separately on a certain time interval [10, 43]. To incorporate these observations into a computational principle, we separate the processing of individual data streams using a self-organizing principle until substantially different sensory information (or a distinctive event) is perceived. Integration of the two sensory modalities takes place only when the timing of a distinctive event encountered by both sensory streams coincides. Since that distinctive event may occur at any moment in time, event-based integration must divide the sensory streams into non-equal time intervals. Event-based integration is a distinctive feature of the proposed integration method.

Synergy in a multimodal integration approach is difficult to quantitatively evaluate. Conflict estimation is a systematic approach that provides guidelines for integration [39, 26, 40]. Conflict studies [39, 26, 40] investigate causes of spatial discrepancies between the shifted spatial layout obtained through vision and the correct spatial layout provided by other sensory modalities, such as proprioception. This discrepancy (conflict) estimation approach has been used to judge spatial direction

perception as measured through target-pointing responses. It can be adapted to evaluate how dynamic visual and non-visual information is integrated over time to determine how distance traveled is perceived while moving.

One approach to integration is to add priority values (weights) to both (visual and proprioceptive) information streams. Priorities are commonly assigned to visual information [39, 26, 40], but the selection criteria are rarely disclosed and there are differing opinions on how modality weights should be determined. One suggestion is to base weights according to the precision of the information in each modality. How the priority values are chosen, however, has not been shown by experimentalists. There are different ideas about how to determine the weights given to each modality. According to one idea, the weights are determined by the precision of the information in each modality [26, 42]; another assigns weights according to the amount of attention directed to each modality [39, 20, 21, 40]. The thinking behind these ideas stems from the concepts underlying statistical optimization models which assume that sensory information from multiple sources should be weighted according to the estimated reliability of each cue. Unlike the discussed issues surrounding weight assignments for vision and proprioception cues in conflict studies [39, 26, 40], we coupled the dynamic view-like and self-motion information using a self-teaching principle, conforming with the application domain of concurrent mapping and navigation [4, 11, 45].

To illustrate our integration approach, two data streams were recorded as an autonomous robot explored an unfamiliar environment. They provided absolute and relative information about the robot's movement with respect to the relation of robot movement to the surrounding objects.

In Sec. 2, we present our integration hypothesis followed by an explanation of the temporal synchronization principle as the framework for the application domain in Sec. 3. The actual integration and the experimental testing is presented in Sec. 4 and the results are discussed in Sec. 5.

2. Hypothesis

Multisensory integration requires an anatomical convergence of unisensory inputs onto a single neuron or ensembles of interconnected neurons [33], and some degree of temporal alignment of the unisensory inputs [36]. Areas associated with multisensory integration include the superior temporal polysensory area, lateral and ventral intraparietal areas. A detailed explication of the brain mechanisms of multisensory processing has been conducted in the carnivore superior colliculus [36] and substantial progress has also been made at the neocortical level, most notably in monkeys [5, 7, 13, 14, 16, 17] and recently in humans [8, 18].

Despite these advances, questions remain about the anatomical substrates of multisensory convergence in primates. Questions also remain about the temporal parameters of the converging sensory inputs. Temporal windows exist for the integration of neural responses to stimulus inputs from different modalities and for

perception of fused multisensory inputs (i.e., relating to the same object [36]). However, only a few studies of the timing of sensory inputs to the neocortex have been reported, focusing mostly on the response latencies in the visual system [26, 32, 30]. Therefore, detailed modeling at this stage remains difficult.

Temporal registration experiments suggest that the brain does not bind information entities from different modalities in real time; instead, it binds the results of its own processing systems on certain time intervals. Our hypothesis provides a constructive basis for an integration strategy. We assumed that the brain operates as a self-organizing information system that processes the sensory flow to asymmetric activation patterns at the separate sensory modality level. Since each modality brings different levels of generalization and information about the external world, information from any one modality can also serve as a "teacher" for other modalities.

Experimental support for this line of thinking comes from studies of the rat head-direction system. Head-direction cells are located in many parts of the rat brain, including the pre- and post-subiculums, the anterodorsal thalamus, and the mammillary nuclei [27, 30, 34, 35]. Head-direction tuning may arise when angular head-velocity signals are fused. These signals originate from vestibular neurons that are tonically active when the head is still, but the firing rate increases when turning is in one direction (i.e., center to left) and decreases when it turns the other way (i.e., center to right).

Experimental evidence also implicates a fast-acting contribution of visual input in the organization of head-direction circuits [6, 8]. That is, the preferred head direction of these cells can be controlled by a visually salient landmark. When the head is rotated, the preferred direction of these cells is generally aligned with the angular displacement of the landmark [37]. These, and other similar experiments [30, 46], suggest a critical role of visual information in the calibration and development of head-direction tuning. Calibration of head-direction cells by visual landmarks has been shown in [44]. The possible role of visual information as a teaching signal that supervises the development of an integrator network has also been studied [15]. They concluded that selective amplification teaches the vestibular input how to predict and replace missing input.

3. Temporal Synchronization in Spatial Navigation Setup

In robots, various sensors asynchronously provide information with different meanings and sampling characteristics. Established ways of combining the information from different information sources are missing. To combine multimodal information sources, the following principles need to be considered:

- Data that are perceived (recorded) at the same time interval relate to the same situation (event).
- Processing of different data streams is done in separate modalities, followed by synchronization using a temporal principle.
- The temporal synchronization is event-based (not fixed-time interval based).

With these guiding principles, the following computational steps will be taken: firstly, event-based time intervals will be defined. Secondly, information recorded within these intervals will be represented so that fusion is possible. Thirdly, the actual integration takes place.

The mapping task is solved by using the data recorded by the robot during its exploration of an unknown environment. Figure 1 shows the experimental environment. Several exploration routes are plotted on the picture. Black points on the floor indicate novel regions that have been clustered in different classes (events) according to the sensory information as detailed below.

Imagine a group of sensors imitating a consummate description of the environment that can be created by biological systems is difficult. A plausible alternative might be to focus on a couple of sensors that provide complementary information about the environment. The relevance of an egocentric perspective for an autonomous robot in spatial modeling of previously unknown environments was elaborated in [4, 45]. That egocentric model combines two types of information: absolute and relative with respect to the relation of robot movement to the surrounding objects. The absolute perspective records sensory information independent of robot movements using laser range finders. Relative information reflects the robot perceptions of its motion. For instance, if the flooring is different, the angular velocity readings may differ when the robot takes the very same trajectory. The relative sensory stream is recorded by a build-in gyroscope.

The "views" that the robot perceives with a laser range finder source the absolute information. Visual information has an absolute character. Since this method takes

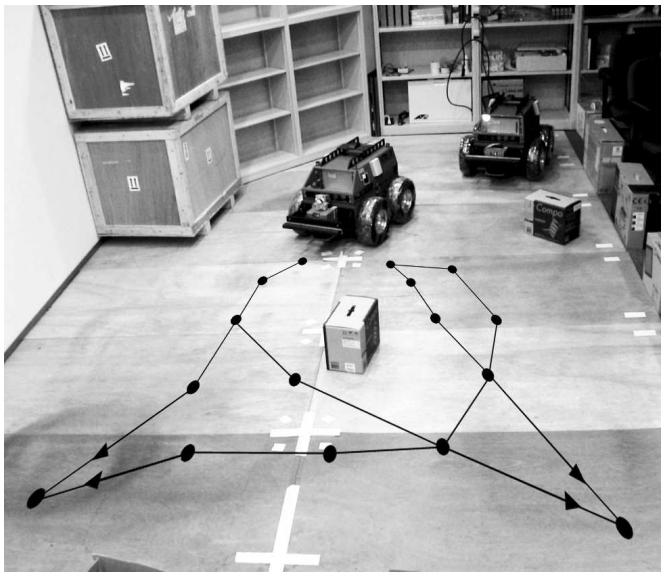


Fig. 1. Experimental environment in the picture several recorded trajectories that reach perceptually similar places have been added.

ideas from the combination of visual and proprioceptive cues by living organisms, the recordings from the laser range finder are referred to as view-like information. Vision is not used because the application domain that the current study is targeting is underwater navigation, an area where visual information is scarcely available. Apparently, the range sensor information is simpler to process and sufficient to represent the idea of fusing absolute with relative egocentric information. The individual "view" of the robot is formed by recording 360 samples per 180 degrees. A snapshot of a polar representation of such a record is shown in Fig. 2.

A sequence of such snapshots, recorded during robot exploration, form a dynamic trajectory:

$$x_i(t_i) = \frac{s_i(t_i) + \sum_{\tau=1}^{h_i-1} f_i(\tau) x_i(t_i - \tau)}{h_i} \tag{3.1}$$

where $s_i(t_i)$ is the readout of sensor (element) i, $t_i > 0$ is an integer time stamp for sensor (element) i, $x_i(t_i)$ specifies the current sensory representation, that keeps a history of $h_i = \min(H_i, t_i)$ elements from the exponentially decaying forgetting curve f. A priori known constant H_i denotes the maximum history length. Items in the dynamic trajectory decay in time, corresponding to the decay theory of forgetting in short term memory. In our experiments a decay kernel [Eq. (3.2)] was used. This is because, of all the previously seen patterns, the last is the most vivid and influences

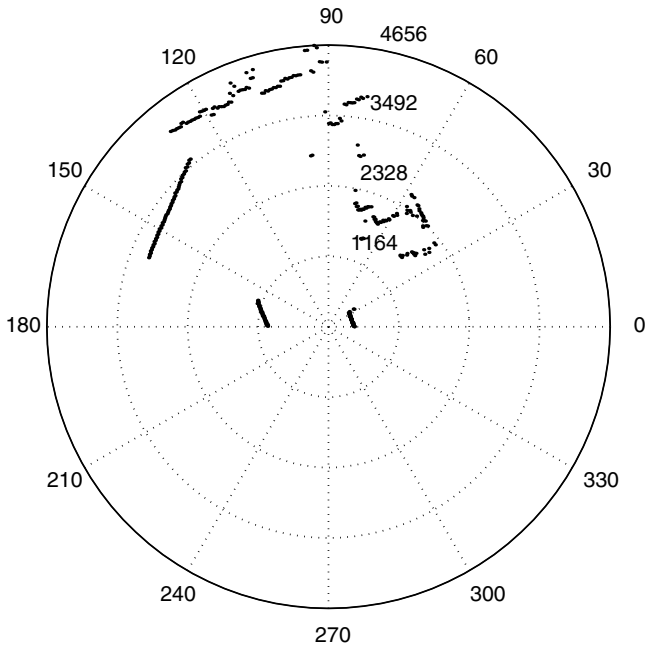


Fig. 2. Sample recording from the laser range finder. The range finder "view" is a composite recording of 360 samples per 180 degrees that forms a snapshot. Snapshots are recorded at frequency of 4.7 Hz. The distances are presented in millimeters.

most strongly the current perception, as the influence of the older patterns fade away.

$$f_i(t) = \exp(-\alpha_i t) \tag{3.2}$$

where constant $\alpha_i \in [0, 1]$ determines the decay profile f_i for sensor (element) i. Each unit in this short-term memory model samples a symbol in a specific time interval. Such a dynamic sequence encodes the first sensory stream, used for the integration (i.e., the laser range finder stream). It represents the absolute perspective of the robot about its own motion (i.e., robot motion with respect to the surrounding objects).

The velocity measured by the gyroscope is used to represent the relative cue that resembles proprioceptive information in animals. It reflects the robot's perception of its own motion. Most informative are the angular velocity recordings from the robot, since they reflect directional changes in its trajectory. The temporal synchronization of these two information streams is performed as follows: a neural gas algorithm [23] is applied to the view-based sensory stream to determine the intervals when a novel "event" occurs. Learning dynamics is guided by a combination of competitive Hebbian learning and vector quantization algorithm. A set of n reference vectors $\vec{w}_i$, $i \in \{1, 2, \ldots, n\}$ have strengths, depending on their neighborhood ranking. When an input vector $\vec{x} = \{x_1(t_1), x_2(t_2), \ldots, x_n(t_n)\}$ is presented, a neighborhood ranking of the reference vectors takes place $(\vec{w}_{i0}, \vec{w}_{i1}, \ldots, \vec{w}_{in})$ with $\vec{w}_{i0}$ being the closest to $\vec{x}$, $\vec{w}_{i1}$ being the second closest to $\vec{x}$, and $\vec{w}_{ik}$, $k \in \{1, 2, \ldots, n-1\}$ is the reference vector for which there are k vectors $\vec{w}_j$ with $\|\vec{x} - \vec{w}_j\| < \|\vec{x} - \vec{w}_{ik}\|$. The rank index associated with $\vec{w}_i$ is denoted by $k_i(\vec{x}, \vec{w})$. Using a Hebbian-like rule, the adaptation step for adjusting is given by:

$$\Delta\vec{w}_i = \varepsilon(t) h_\lambda(k_i(\vec{x}, \vec{w}))(\vec{x} - \vec{w}_i) \tag{3.3}$$

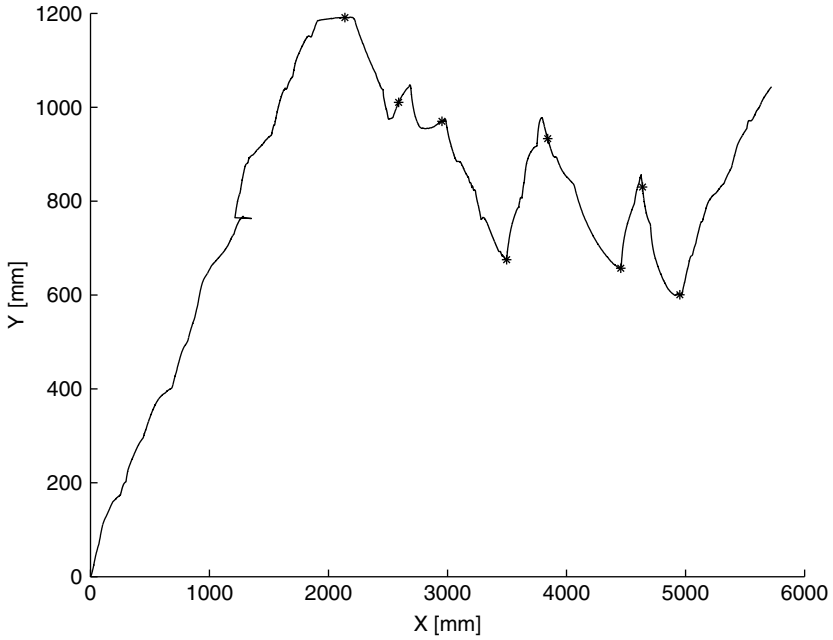
where the step size $\varepsilon \in [0, 1]$ is the learning rate, and $h_\lambda(k_i(\vec{x}, \vec{w})) \in [0, 1]$ accounts for the topological arrangement of the $\vec{w}_i$ in the input space.

$$h_\lambda = \exp\left(-\frac{k}{\lambda}\right) \tag{3.4}$$

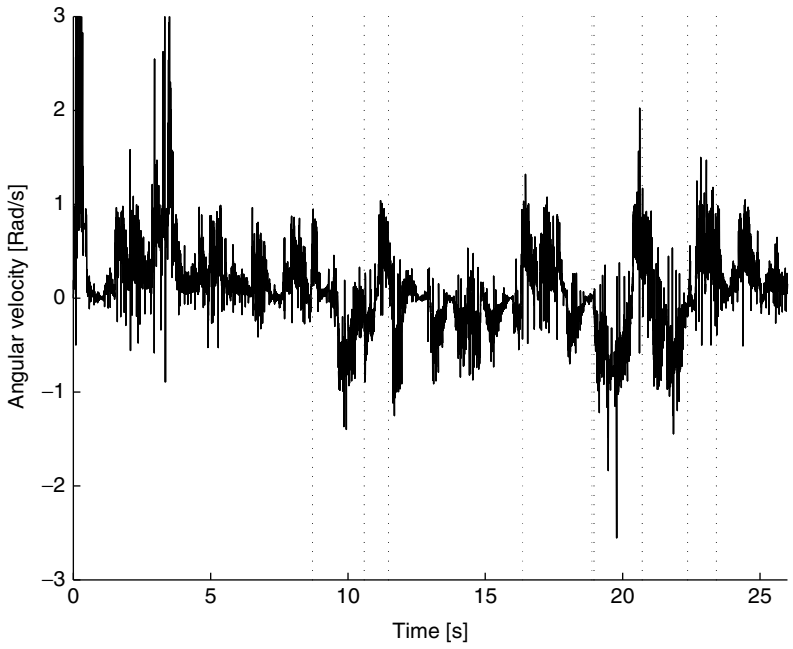
i.e., the neighborhood relies on the rank in the ordered sequence of distances, and the weights are learned according to (3.3), while decreasing λ. For the simulation, the $\varepsilon(t)$ and $\lambda(t)$ are calculated as follows:

$$g(t) = g_0 \left(\frac{N}{g_0}\right)^{\frac{t}{T}} \tag{3.5}$$

where $g \in \{\lambda, \varepsilon\}$, $N = 0.01$, $\varepsilon_0 = 0.5$, $\lambda_0 = n/2$, where n is the number of neurons, and T is the number of the training patterns. The simulations were made with $n = 20$ neurons, sufficient to encode the different patterns from the experimental environment. In order to extend the method to any environment, an incremental version of the algorithm known as a growing neural gas algorithm [12] can be used.



(a)



(b)

Fig. 3. Temporal synchronization principle. (a) Experimental trajectory is segmented by the clustering algorithm. Points on the curve denote the events that are determined as new. (b) The corresponding angular velocity curve, segmented according to the time the robot has spent in the same event cluster.

This algorithm starts from two neurons and increases the neurons number, until all the different patterns are encoded to separate classes. A Euclidean distance measure decides how many classes to form.

Synchronously, an event-based segmentation is performed on the second (velocity) sensory stream. Note that the dynamic events are formed on-line with minimal processing of incoming data.

Figure 3 illustrates the temporal synchronization of the two information streams. Figure 3a represents a trajectory that the robot took during its exploration of an environment. Any qualitatively different "view" that the robot observes is defined as new segment in the environment and it is denoted by * in Fig. 3. The duration of the trajectory in these segments determines the division of the other sensory data stream as illustrated in Fig. 3b.

The results of a temporal synchronization by a recorded short experimental trajectory are shown in Fig. 4. Figure 4a shows view-based segmentation of a two-dimensional space, obtained during free exploration of the robot. The points where a qualitatively new view occurs, as determined by the described algorithm, are shown. The robot trajectory is not shown on the plot. Corresponding velocity trajectory is reconstructed by angular and linear velocity recordings (Fig. 4b). Figure 4c shows the velocity curve after temporal synchronization with the view clusters.

The synchronization process is as follows: after clusters of the view based information stream are found, the velocity data are segmented on the same time intervals, considering the different sampling frequencies of both sensory streams. Based on this segmentation, a clustering to unified trajectory elements (or motion primitives) of the velocity curve is made, using the algorithm described in Eqs. (3.3)–(3.5). The input vectors by this clustering are the segments from the velocity curve. Figures 4d–f depict the resulting motion primitives corresponding to the shown segments of the velocity curve.

4. Integration Results

In the learning phase, the sequences of views were clustered by the self-organizing algorithm, forming dynamic events throughout the robot's continuous exploration of the unknown environment. Those moments when a qualitatively new view is recorded by the gyroscope become the temporal dividers for the proprioceptive dynamic sensory stream. Similar steps in the testing phase were performed with the sensory data gathered from further exploration. Specific trajectories were recorded for the testing process where classification, not clustering, is done based on the clustering categories already attained from the exploration data. The database of distinctive events is built gradually during on-line implementation. Figure 5 summarizes the computational flow by off-line event based integration. A supervised neural network is used for teaching between the modalities, that are denoted as multimodal integration blocks in the figure.

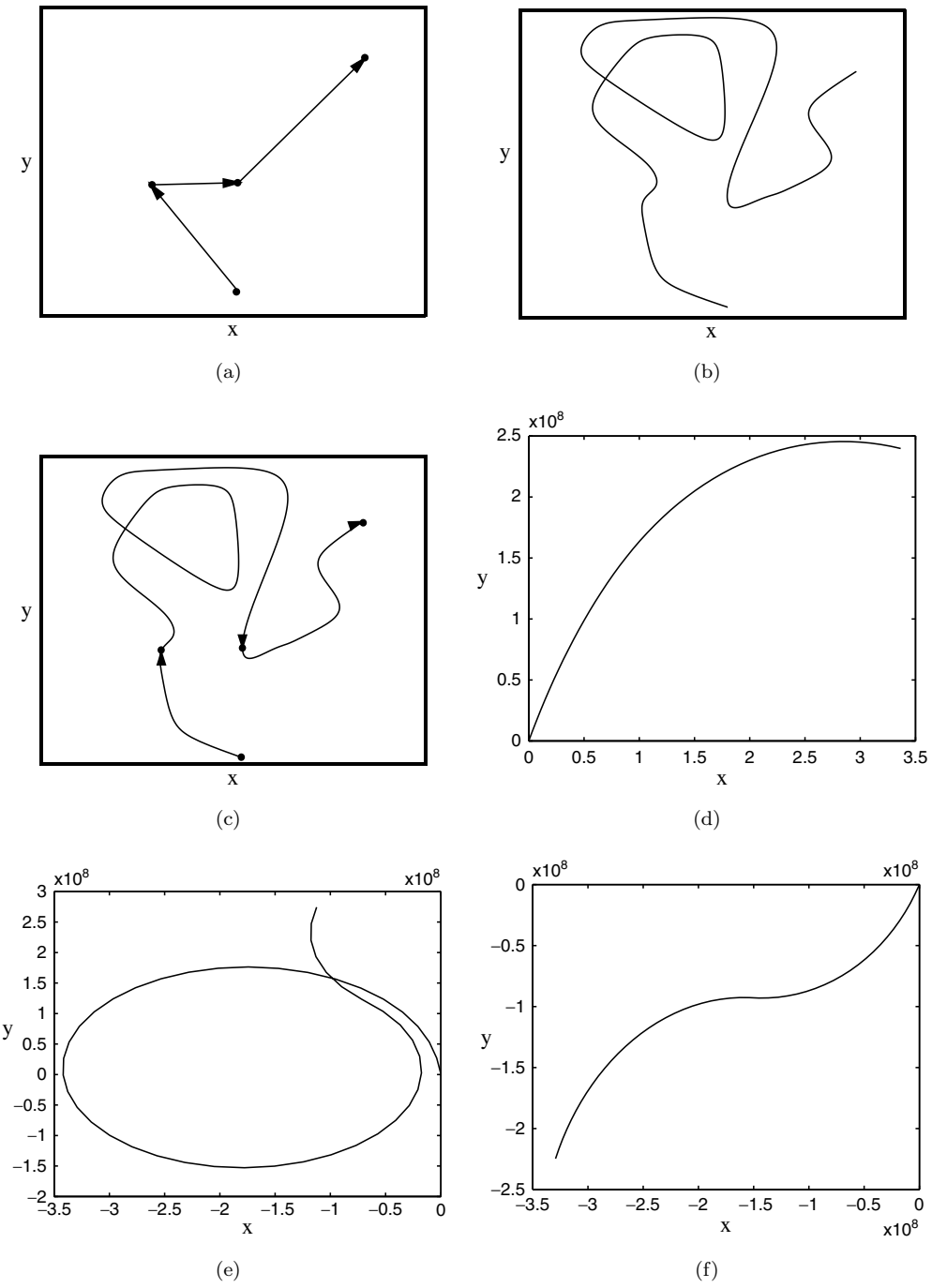


Fig. 4. The result of temporal synchronization over a short trajectory recorded by the robot. (a) View-based segmentation of a two dimensional space. (b) The corresponding velocity trajectory (c) Temporally synchronized velocity curve. (d–f) Classes, corresponding to the segments of the velocity curve.

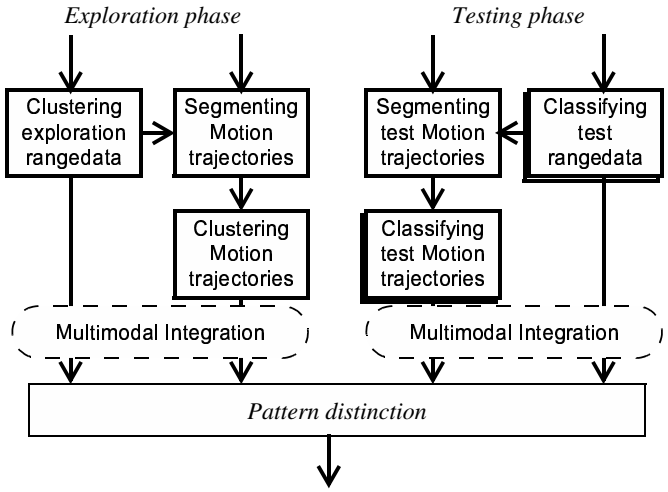


Fig. 5. Information flow of event-based integration, consisting of exploration and testing phases that reflect the experimental process.

Several groups of experiments were performed according to this computational scheme. In the first group of experiments, perceptually similar places with respect to the view-based sensory stream were training inputs and motion primitives perceived by the auto-motion (velocity) stream functioned as teacher. Figure 6 depicts the dynamic trajectories, obtained from the view-based information stream.

All trajectories reach one of two corners in the experimental environment, which look the same if "seen" by the laser range sensors (Fig. 6a). Figures 6b–f depict the encoding by dynamic trajectory formation. The plots on the left show the dynamic trajectory and the plots on the right show the classes of distinctive views encountered while making these trajectories. Considering only the temporal history of experienced views, the dynamic trajectory method can distinguish most of the trajectories, although two, shown in Figs. 6e–f, appear similar despite the obviously different view history. Using the self-motion primitives (that can be obtained from the velocity sensory cue) as a teaching signal, all dynamic trajectories are disambiguated. Experimental analysis shows that using self motion as a teaching signal helps perceptual aliasing, but sometimes distinguishes trajectories that are very similar.

The view-based signal was used as teacher and the self motion signal as an input in the second group of experiments. New motion trajectories, specifically selected for their similarity with respect to classification to motion primitives, were trained. The obtained results coincided better with decisions that a human observer would make.

The later experiments were made with trajectories that were not controlled with respect to similarities. Correspondingly, short, long, and arbitrary length sequences were recorded. The environment contains 6 big objects that divide it on routes that can be distinguished by a human observer. All possible trajectories between

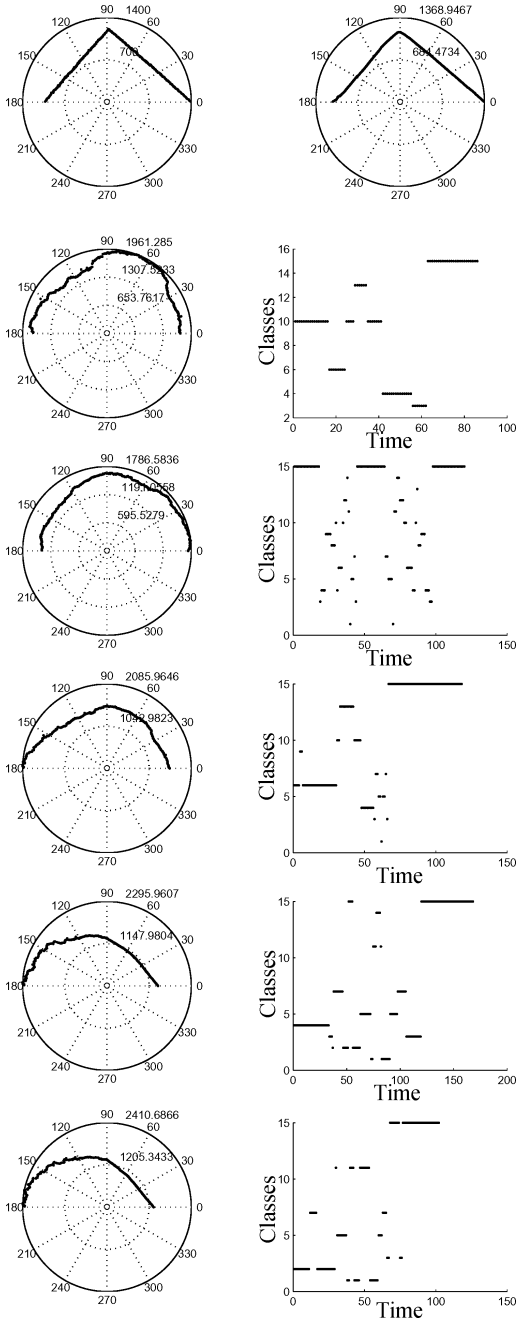


Fig. 6. Environmental similarities and the corresponding dynamic trajectories. (a) The left plot shows the laser range finder recording of a corner in the experimental room. The plot on the right shows the network output. (b–f) 5 dynamic trajectories that finish at that corner (i.e., have the same final sensory reading). Alternative routes to the corner are shown as clusters in the right plots.

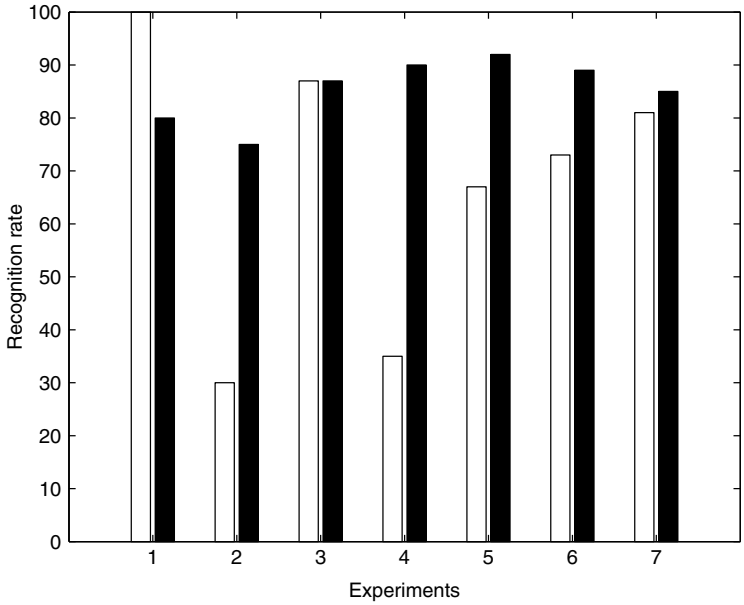


Fig. 7. Recognition rate with view based information used as a teacher (black bars) and self-motion information used as a teacher (white bars). The experiments are made as follows: 1-perceptually similar trajectories with respect to view based sensors; 2-perceptually similar trajectories as perceived by self-motion sensors; 3-short trajectories; 4-long trajectories; 5–7 non selected trajectories.

the same objects or object traversed in the same order belong to the same route. For the learning phase the robot has passed all the possible routes between the environmental objects at least once. The test sequences were recorded by a robot traversing trajectories within the possible routes. Every trajectory record consists of view sequence of data and velocity sequences for reconstructing the self-motion trajectories. The recognition of a trajectory was evaluated manually.

Experimental results where self-motion signals were used to teach are shown in Fig. 7 as white bars. Results with the view signal as teacher are plotted with black bars in the same figure. The recognition rate was evaluated as the percentage of the trajectories from the given group that was classified in the right class.

Classification over the same training sets was also made with the weighting integration method, implemented after [39]. The bigger weights are assigned to view information; however, better results with current method, as shown in Fig. 8, are not achieved even after thorough experimental testing of both algorithms. They are validated for this particular case only.

5. Discussion

The proposed multimodal integration method provides an alternative way to combine information from different sensory modalities. It addresses two groups of problems — firstly, how and when the combination between the two sensory streams

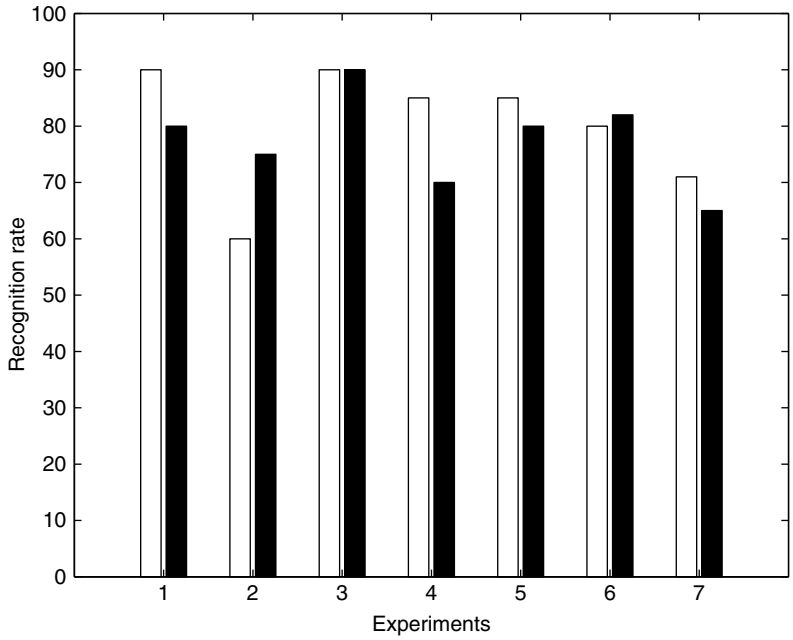


Fig. 8. Recognition rate by weighted integration (black bars) and integration with a teacher (white bars). The experiments are made as follows: 1-perceptually similar trajectories with respect to view based sensors; 2-perceptually similar trajectories as perceived by self-motion sensors; 3-short trajectories; 4-long trajectories; 5–7 non selected trajectories.

occurs, and secondly, how synergy in the integration process is obtained. The established synchronization framework is an efficient way to bring the multimodal information together at event-based temporal intervals. This approach prevents error accumulation since the synchronization takes place over relatively short temporal intervals (i.e., the accumulative errors are reset after every interval). Another advantage of this model is that both the consequent perceptions as well as the transitions between them are dynamically encoded.

Information from both sensory streams is used directly, without feature extraction. The basis for this synergistic integration is tested for data sets recorded on the following principles: perceptual similarity of the trajectories with respect to view-based information; perceptual similarity of the trajectories as perceived by self-motion sensors; different length of the trajectories. The experimental results show that in most cases using a view-based sensory stream as a teacher is advantageous. Trajectories, recorded on a visual similarity principle can be successfully disambiguated if the self-motion (velocity) signal is used as a teacher, but this disambiguation is crude, occasionally not differentiating between similar trajectories. Disambiguation ensured by the view-based teacher never reaches 100 percent recognition of experienced trajectories, but gives results that most closely mimic human observer decision behavior.

This integration method is compared to a weighted integration method that uses visual and proprioceptive sensory cues (see [39]). This is possible since two sensory streams are used: view-based information resembles animal vision, and velocity information is equivalent to proprioception by living organisms in this experimental setting. In most of the testing set groups, the integration process gives better results by the method, as proposed in this paper, except when there are similarities in self-motion sensory stream or when the data sets are absent.

From the third and forth groups' test sets we can conclude that the proposed integration method explains perceptual hierarchy in the following way: knowledge about instant movement contributes significantly to short-term navigation, while visual perceptions have bigger impacts over longer terms.

While temporal synchronization can be implemented as an on-line learning process, the multimodal integration requires off-line processing. This is one major drawback of the model.

We base our multimodal integration on studies that rely on visual and proprioceptive data streams. Such streams have different dimensionality. However, the experiments show integration of two unidimensional data streams. Using visual information is a straightforward extension to the developed integration method, since the two integration streams are processed separately. The used algorithm can work with two-dimensional data as well. The speed of processing, however, will be a problem for a real robotic implementation. To deal with that, we have the following idea. The range sensor readings give a sparse depth representation. The same could be obtained by using unimodal vision features. For instance the most salient features of corners/crossing or edges could be used to construct a depth map [22].

Further, we plan to model the integration process on the level of cortical structures, including temporal synchronization with feedback, since recent work suggests that feedforward and feedback projections contribute to the convergence of the integrated representation. We will use visual and proprioceptive cues.

Acknowledgments

I (E.B.) would like to acknowledge the multiple discussions with U.R. Zimmer (Australian National University, Canberra) for the multiple discussions. We have done some initial experimental work together, see [4].

References

[1] Andersen RA, Multimodal integration for the representation of space in the posterior parietal cortex, *Philos T Rly Soc B* **352**(1360):1421–1428, 1997.

[2] Babeau V, Gaussier P, Joulain C, Revel A, Banquet J, Merging visual place recognition and path integration for cognitive map learning, *SAB*, 101–110, 2000.

[3] Banquet J, Gaussier P, Quoy M, Revel A, From reflex to planning: Multimodal, versatile, complex systems in biorobotics, *Behav Brain Sci* **24**(6):1051–1053, 2001.

[4] Barakova EI, Zimmer UR, Global spatial modeling based on dynamics identification according to discriminated static sensations, in *Proceedings of Underwater Technology*, IEEE, Tokyo, 2000.

[5] Benevento LA, Fallon J, Davis BJ, Rezak M, Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey, *Exp Neurol* **57**:849–872, 1977.

[6] Blair H, Sharp P, Anticipatory head direction signals in anterior thalamus: evidence for a thalamocortical circuit that integrates angular head motion to compute head direction, *J Neurosci* **15**:6260–6270, 1995.

[7] Bruce C, Desimone R, Gross CG, Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque, *J Neurophysiol* **46**:369–384, 1981.

[8] Canon LK, Directed attention and maladaptive "adaptation" to displacement of the visual field, *J Exp Psychol* **88**:403–408, 1971.

[9] Chance SS, Gaunet F, Beall AC, Loomis JC, Locomotion mode affects the updating of objects encountered during travel: the contribution of vestibular and proprioceptive inputs to path integration, *Presence-Teleop Virt* **7**:168–178, 1998.

[10] Eagleman DM, Sejnowski TJ, Motion integration and postdiction in visual awareness, *Science* **287**(5460):2036–2038, 2000.

[11] Elfes A, Sonar-based real-world mapping and navigation, *IEEE J Robotic Autom* **RA-3**(3):249–265, June 1987.

[12] Fritzke B, A growing neural gas network learns topologies, *Adv Neural Inf Process Syst* 625–632, 1995.

[13] Graziano MS, Hu XT, Gross CG, Visuospatial properties of ventral premotor cortex, *J Neurophysiol* **77**:2268–2292, 1997.

[14] Graziano MS, Yap GS, Gross CG, Coding of visual space by premotor neurons, *Science* **266**:1054–1057, 1994.

[15] Hahnloser RHR, Emergence of neural integration in the head-direction system by visual supervision, *Neuroscience* **120**(3):877–891, 2003.

[16] Hikosaka K, Iwai E, Saito H, Tanaka K, Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey, *J Neurophysiol* **60**:1615–1637, 1988.

[17] Hyvarinen J, Shelepin Y, Distribution of visual and somatic functions in the parietal associative area 7 of the monkey, *Brain Res* **169**:561–564, 1979.

[18] Jousmaki V, Hari R, Parchment-skin illusion: sound-biased touch, *Curr Biol* **8**(6):R190, 1998.

[19] Kam M, Zhu X, Kalata P, Sensor fusion for mobile robot navigation, *P IEEE* **85**(1), 1997.

[20] Kelso JAS, Cook E, Olson ME, Epstein W, Allocation of attention and the locus of adaptation to displaced vision, *J Exp Psychol Human* **1**:237–245, 1975.

[21] Loomis JM, Da Silva JA, Fujita N, Fukusima SS, Visual space perception and visually directed action, *J Exp Psychol Human* **18**:906–921, 1992.

[22] Lourens T, Würtz RP, Extraction and matching of symbolic contour graphs, *Int J Pattern Recogn* **17**(7):1279–1302, 2003.

[23] Martinetz TM, Berkovich SG, Schulten KJ, Neural gas network for vector quantization and its application to time-series prediction, *IEEE T Neural Networ* **4**:558–569, 1993.

[24] Menzel R, Geiger K, Chittka L, Joerges J, Kunze J, Ler U, The knowledge base of bee navigation, *J Exp Biol* **199**(Pt 1):141–146, 1996.

[25] Mittelstaedt ML, Mittelstaedt H, Idiothetic navigation in humans: estimation of path length, *Exp Brain Res* **139**:318–332, 2001.

[26] Pick HL, Warren DH, Hay JC, Sensory conflict in judgments of spatial direction, *Percept Psychophys* **6**:203–205, 1969.

[27] Ranck JJ, Head-direction cells in the deep cell layers of the dorsal presubiculum in freely-moving rats, *Soc Neurosci* **10**:599, 1984.

[28] Rieser JJ, Ashmead DH, Talor CR, Youngquist GA, Visual perception and the guidance of locomotion without vision to previously seen targets, *Perception* **19**:675–689, 1990.

[29] Roumeliotis S, Bekey G, An extended kalman filter for frequent local and infrequent global sensor data fusion, *Proc SPIE* **3209**, 1997.

[30] Save E, Poucet B, Involvement of the hippocampus and associative parietal cortex in the use of proximal and distal landmarks for navigation, *Behav Brain Res* **109**(2):195–206, 2000.

[31] Schlack A, Hoffmann KP, Bremmer F, Interaction of linear vestibular and visual stimulation in the macaque ventral intraparietal area (vip), *J Neurosci* **21**:23–29, 2001.

[32] Schmolesky M, Wang Y, Hanes DP, Thompson KG, Leutgeb S, Schall JD, Leventhal AD, Signal timing across the macaque visual system, *J Neurophysiol* **79**:3272–3278, 1998.

[33] Schroeder CE, Mehta AD, Foxe JJ, Determinants and mechanisms of attentional modulation of neural processing, *Front Biosci* **6**:672–684, 2001.

[34] Seltzer B, Pandya DN, Afferent cortical connections and architechtonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey, *Brain Res* **149**:1–24, 1978.

[35] Seth AK, McKinstry JL, Edelman GM, Krichmar JL, Visual binding through reentrant connectivity and dynamic synchronization in a brain-based device, *Cereb Cortex* **14**:1185–1199, 2004.

[36] Stein BE, Meredith MA, *The Merging of the Senses*, MIT Press, Cambridge, MA, 1993.

[37] Taube J, Muller R, Ranck J, Head-direction cells recorded from the postsubiculum in freely moving rats: Ii. the effects of environmental manipulations, *J Neurosci* **10**:436–447, 1990.

[38] Tononi G, Sporns O, Edelman GM, Reentry and the problem of integrating multiple cortical areas: Simulation of dynamic integration in the visual system, *Cereb Cortex* **2**:310–335, 1992.

[39] Van Beers RJ, Sittig AC, Denier van der Gon JJ, Integration of proprioceptive and visual position-information: an experimentally supported model, *J Neurophysiol* **81**:1355–1364, 1999.

[40] Warren DH, Schmitt TL, On the plasticity of visual-proprioceptive bias effects, *J Exp Psychol Human* **4**:302–310, 1978.

[41] Wehner R, Michel B, Antonsen P, Visual navigation in insects: Coupling of egocentric and geocentric information, *J Exp Biol* **199**(1):129–140, 1996.

[42] Welch RB, Widawski MH, Harrington J, Warren DH, An examination of the relationship between visual capture and prism adaptation, *Percept Psychophys* **25**:126–132, 1979.

[43] Zeki S, Functional specialization in the visual cortex of the rhesus monkey, *Nature* **274**:423–428, 1978.

[44] Zhang K, Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory, *J Neurosci* **16**:2112–2126, 1996.

[45] Zimmer UR, Robust world-modelling and navigation in a real world, *Neurocomputing* **13**:247–260, 1996.

[46] Zugaro M, Berthoz A, Wiener S, Background, but not foreground, spatial cues are taken as references for head-direction responses by rat anterodorsal thalamic neurons, *J Neurosci* **21 RC145**:1–5, 2001.