CrossMark

Gradual formation of visual working memory representations of motion directions

Hiroyuki Tsuda[1] · Jun Saiki[1]

Abstract
Although it is well accepted that the formation of visual working memory (VWM) representations from simple static features is a rapid and effortless process that completes within several hundred milliseconds, the storage of motion information in VWM within that time scale can be challenging due to the limited processing capacity of the visual system. Memory formation can also be demanding especially when motion stimuli are visually complex. Here, we investigated whether the formation of VWM representations of motion direction is more gradual than that of static orientation and examined the effects of stimulus complexity on that process. To address these issues, we examined how the number and the precision of stored items in VWM develop over time by using a continuous report procedure. Results showed that while a viewing duration of several seconds was required for the successful storage of multiple motion directions in VWM regardless of motion complexity, the accumulation of memory precision was much slower when the motion stimulus was visually complex (Experiments 1 & 2). Additional experiments showed that the difference in memory performance for simple and complex motion stimuli cannot be explained by differences in signal-to-noise levels of the stimulus (Experiment 3). These results demonstrate remarkable temporal limitations in the formation of VWM representations for dynamic objects, and further show how this process is affected by stimulus properties such as visual complexity and signal-to-noise levels.

Keywords Visual working memory · Motion perception · Consolidation · Complexity · Noise

Visual working memory (VWM) functions as an interface between the mind and the visual world by providing a temporary storage for the characteristics of our surroundings. Although the storage capacity of VWM is limited to only three or four objects (Luck & Vogel, 1997, 2013), our visual system functions well during most of our daily activities, and we are mostly unaware of this memory limitation (e.g., "change blindness blindness"; Levin, Momen, Drivdahl, & Simons, 2000). In theory, a large buffering capacity is not necessary as long as the outside world itself can serve as an external memory that can be accessed just by looking at objects whenever needed (Ballard, Hayhoe, & Pelz, 1995; O'Regan, 1992). However, for this type of scheme to be successful, the information transfer from external to internal representations needs to be rapidly completed. Indeed, previous studies have provided abundant evidence that the encoding and consolidation of representations in VWM takes place very quickly, at least for simple, static features such as color or orientation (see Ricker, 2015, for a review). However, the objects and agents that we encounter in everyday life are often dynamic, and further tend to be visually more complex than colored patches or oriented bars (Orhan & Jacobs, 2014). Observers require vision with scrutiny for the precise and detailed processing of such stimuli, which occurs at later stages of visual processing (Hochstein & Ahissar, 2002). Therefore, the immediate access and storage of visual information and its details may not necessarily hold true for dynamic objects, but there have been few studies on the functioning of VWM for motion stimuli so far.

It is well documented that the encoding and consolidation of visual information into VWM operates quickly and efficiently for static visual features such as color and orientation. Converging evidence suggests that both processes are

✉ Hiroyuki Tsuda
  tsuda@cv.jinkan.kyoto-u.ac.jp

✉ Jun Saiki
  saiki.jun.8e@kyoto-u.ac.jp

[1] Graduate School of Human and Environmental Studies, Kyoto University, Yoshida-nihonmatsu-cho, Sakyo-ku, Kyoto 606-8501, Japan

completed within about 100 to 200 ms (Jolicoeur & Dell'Acqua, 1998; Pinto, Sligte, Shapiro, & Lamme, 2013; Ricker, 2015; Sligte, Scholte, & Lamme, 2008; Vogel, Woodman, & Luck, 2006). A recent study has also found that the consolidation of motion direction in VWM can be achieved within this time scale (Rideaux, Apthorp, & Edwards, 2015). In this study, subjects were presented with moving dots followed by backward masking, and the minimum presentation duration for the consolidation of the presented stimuli was determined by a staircase procedure. The average consolidation duration was 82 ms, which was somewhat longer than the durations reported for color and orientation (about 60 ms; Becker, Miller, & Liu, 2013; Mance, Becker, & Liu, 2012), but well within the order of the typical time scale of VWM consolidation, suggesting that the encoding and consolidation of motion does not necessarily require particularly long periods of time. However, the study was primarily interested in whether VWM can consolidate multiple items simultaneously or not, and the aforementioned duration basically reflects the processing time for storing the minimum information that was necessary to perform the task. In addition, the task required only coarse discrimination of the motion direction between four diagonal directions. In contrast, in the current study, we aimed to examine how detailed representations of motion information are constructed in VWM, with a special interest in its time course, in order to describe more comprehensively the formation of VWM for motion information. Considering that the creation of a detailed VWM representation requires more time and dedicated information processing than the creation of a coarse discrimination (Gao, Ding, Yang, Liang, & Shui, 2013; Gao, Gao, Li, Sun, & Shen, 2011; Xu & Chun, 2009), a significant increase in processing time would be required in situations of this kind.

Encoding of motion information in VWM is demanding, especially when the stimulus is visually complex. An important class of complex motion stimuli that we encounter in everyday life is the biological motion (BM) of human actions. Previous studies have investigated the capacity of VWM for storing the BM of observed actions (Shen, Gao, Ding, Zhou, & Huang, 2014; Smyth, Pearson, & Pendleton, 1988; Wood, 2007) or gait patterns (Poom, 2012) from the point-light display stimuli (Johansson, 1973). In those studies, the presentation duration of the to-be-recalled items ranged from 500 ms to 5 seconds, which was much longer than those in typical visual working memory tasks that used simple static features (100 to 300 ms; e.g., Luck & Vogel, 1997). This seems to reflect the need for a longer processing time (presumably for both perceptual and memory formation stages) in the case of dynamic objects that are visually complex. In fact, Shen et al. (2014) showed that at least more than 1 second was required to encode multiple (more than three, simultaneously presented) BM stimuli. It can be argued that the impairment of VWM performance for BM stimuli with short presentation durations

may reflect the decrease in the number of stored items (i.e., failures in encoding or maintenance), but can also be the consequence of the decreased precision (resolution) of stored representations, although both possibilities may not be mutually exclusive. Since the previous studies were conducted with change detection paradigms, it was difficult to separately estimate the capacity and the precision of stored items in VWM, making it challenging to specify the difficulty of forming VWM representations from complex motion stimuli. It would be useful to determine how each aspect of memory performance (capacity and precision) develops over time, but the temporal aspect of creating VWM representations for motion information has been rarely examined.

Consequently, the aim of the present study was to examine the temporal characteristics of VWM performance for motion stimuli in terms of both the number and the precision of stored representations, and to examine how it is affected by stimulus complexity. We employed two types of motion stimuli, biological motion of a point-light walker representing complex motion, and a random dot kinematogram (RDK) representing simple motion. The RDK was considered as simple motion because it consists of homogeneous motion signals, whereas BM is complex because it consists of dots with different spatiotemporal energies and requires the higher-order integration of the local motion signals to acquire a coherent percept of its walking direction (Blake & Shiffrar, 2007). We employed a continuous report procedure (e.g., Zhang & Luck, 2008) to acquire separate estimates of capacity and precision of representations in VWM, in which participants were asked to recall the walking direction of the BM of a point-light walker, or the motion direction of an RDK. We examined how the viewing duration and the set size of the sample array affected VWM performance. Based on previous VWM studies that used BM stimuli (Poom, 2012; Shen et al., 2014; Smyth et al., 1988; Wood, 2007) or complex static objects (Alvarez & Cavanagh, 2004; Curby & Gauthier, 2007; Eng, Chen, & Jiang, 2005), the presentation duration of the sample array in the current study spanned from 500 ms to 5 seconds, which was expected to effectively capture the temporal profile of creating VWM representations from complex motion stimuli.

In the following study, we first conducted experiments by using BM as a stimulus of interest (Experiment 1). Then, Experiment 2 was conducted to examine how performance may differ for different types of stimuli: simple motion (RDK), complex motion (BM), and static orientation (as a control condition). If VWM representations of complex motion form slower than those of simple motion, it will be reflected in the number and/or the precision of stored items. Finally, Experiment 3 was conducted as an additional control experiment in order to address a potential alternative account for the effect of stimulus complexity on VWM performance (see the corresponding section for details).

Experiment 1a

We first investigated the temporal development of VWM representations of motion direction by using the biological motion (BM) of a point-light walker. Participants were asked to recall the heading direction (walking direction) of the BM by adjusting a probe stimulus. The effects of viewing duration (500; 1,000; 2,500; and 5,000 ms) and the set size (the number of items in the sample array, one and five) of the sample array on memory performance were tested. The viewing duration was randomly varied across trials, and the different set sizes were tested in separate blocks of trials.

Method

Participants Eighteen Kyoto University students (18–22 years old) participated in the experiment for monetary compensation. All participants reported normal color vision and had normal or corrected-to-normal visual acuity. Each participant gave informed consent to participating in the experiment. All procedures were preapproved by the Ethics Committee of Kyoto University. The sample size was specified so that it is comparable to or larger than that of previous research that used a similar stimulus or experimental procedure (Bays, Gorgoraptis, Wee, Marshall, & Husain, 2011; Poom, 2012; Shen et al., 2014).

Stimuli and apparatus The motion of a point-light walker was adapted from a data set by Vanrie and Verfaillie (2004). The walker was made up of 13 white dots with a diameter of 0.12°, and the walker's overall height was 4.1°. The walking motion was one gait cycle per second, with a sampling rate of 60 Hz, which was resampled from the original 30 Hz data with linear interpolation. The walking directions of walkers in the sample array were randomly assigned between 0° and 360° (see Fig. 1 for an example display). The backward mask was a phase-scrambled and location-scrambled version of the original walker, which is most effective for the masking of BM (Bertenthal & Pinto, 1994; Thurman & Grossman, 2008). The stimuli were displayed on a CRT monitor with a 120-Hz refresh rate, using the software Processing (http://processing.org). Participants were seated approximately 57 cm from the monitor in a dark room, with their heads stabilized by a chin rest.

Design and procedure Each trial began with the presentation of a central fixation cross (white, 0.62° diameter) for 500 ms against a dark gray background (Fig. 1; see also the movie in the online Supplementary Material). A memory array was then presented, consisting of one or five items. Stimuli were displayed at evenly distributed positions on an invisible circle with a radius of 6.5°. The memory array was presented for 500; 1,000; 2,500, or 5,000 ms, followed by the mask for 200

ms. Participants were asked to remember the heading directions of the walkers. After a blank period of 1,000 ms, a probe item was presented at one randomly chosen location from the preceding memory array. Participants adjusted the heading direction of the probe walker to match that of the walker that had been presented at the same location in the preceding memory array. Moving the mouse leftward or rightward rotated the probe walker. Participants were instructed to respond as accurately as possible, and no time limit was imposed.

The experiment was a 4 (viewing duration: 500; 1,000; 2,500; or 5,000 ms) × 2 (set size: one or five items) within-subjects design. There were 52 trials in each condition, yielding 416 trials in total. The presentation duration was randomly varied across trials, and the different set sizes were tested in separate blocks of trials. There were eight blocks in total, and participants were allowed to rest at the end of each block. Before the experimental trials, participants were given practice trials and familiarized themselves with the procedure.

Analysis For each trial, we calculated the error of the recalled direction by subtracting the correct value from the participant's response. Note that walkers facing toward and away from the observer were visually almost identical and indistinguishable—that is, if a walker had a walking direction of $N°$, then it is indistinguishable from a walker whose walking direction is $(180 − N)°$, where walking direction is defined as 0° when the walker is facing directly toward the observer and ±90° when it is facing rightward or leftward (orthogonal to the observer's line of sight). Therefore, although responses were coded in the range −180° to 180°, the range of unique directions was −90° to 90°, so that we first remapped the data accordingly. The distribution of errors was fit to the mixture model put forth by Bays, Catalao, and Husain (2009) to evaluate (1) the standard deviation (SD) of the von Mises distribution, which corresponds to memory precision; (2) the probability of random guesses (guess rate); and (3) the swap rate or the probability of nontarget responses, which are caused by misreporting a nontarget value. The hypothesis regarding the effects of the experimental manipulations was tested by a two-factor repeated-measures ANOVA. Significant main effects were followed by Shaffer's modified sequentially rejective Bonferroni procedure (the method improves power while controlling type one error; Shaffer, 1986).

Results and discussion

The effects of viewing duration and set size on the SD, guess rate, and swap rate are summarized in Fig. 2.

Memory precision gradually improved as the viewing duration increased, which was reflected in the monotonous decrease in the SD. There were main effects of both duration, $F(3, 51) = 7.10$, $p < .001$, $\eta_p^2 = .29$, and set size, $F(1, 17) = 265$, $p < .001$, $\eta_p^2 = .94$, on the SD, but the interaction was not
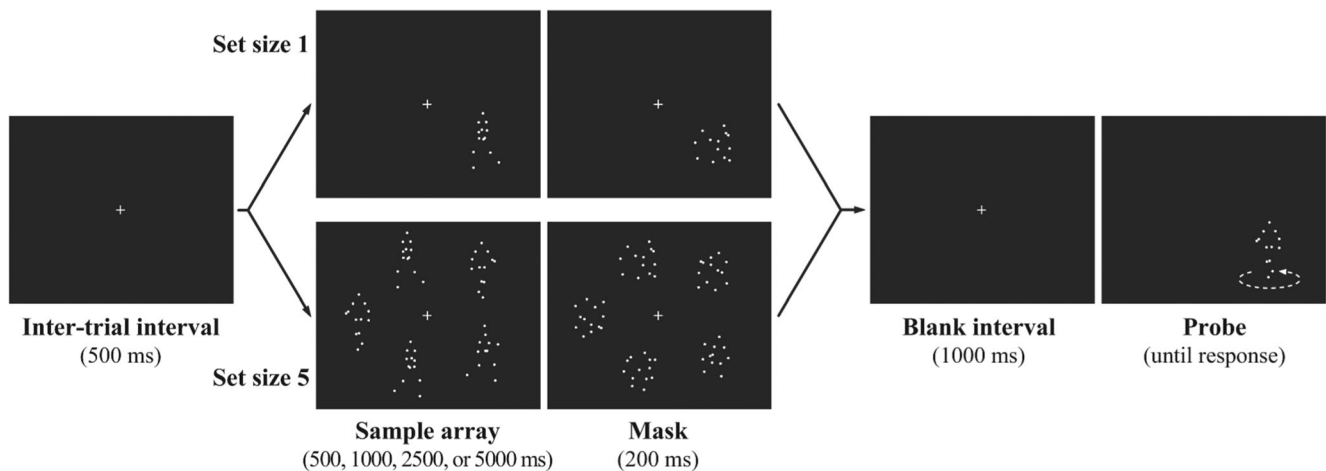
Fig. 1 Schematic illustration of a single trial in Experiment 1a. Participants were presented with one or five point-light walkers for variable durations (500; 1,000;, 2,500; or 5,000 ms) followed by motion masks for 200 ms. After a blank period of 1 s, a probe item was presented, and participants adjusted the heading direction of the probe walker to match that of the item that had been presented at the same location in the sample array. The circling arrow in the right-most panel is shown for illustrative purposes only and was not actually presented

significant, $F(3, 51) = 2.12$, $p = .11$, $\eta_p^2 = .11$. The result indicates that the quality of memory representation improved with increasing viewing duration up to several seconds, and this was true even when the set size was one, which suggests that the development of memory precision for a BM stimulus was much slower than that for simple features such as orientation and color.

The number of items stored in memory also increased with longer viewing durations, as this was reflected in the decline in the guess rate. There were main effects of both duration, $F(3, 51) = 3.92$, $p = .014$, $\eta_p^2 = .19$, and set size, $F(1, 17) = 33.1$, $p < .001$, $\eta_p^2 = .67$, on the guess rate, and the interaction was also significant, $F(3, 51) = 4.55$, $p = .0067$, $\eta_p^2 = .21$. Interestingly, the guess rate reached nearly zero (2.7%) in the Set Size 5 condition when viewing duration was 5,000 ms, which did not differ from that observed in the Set Size 1 condition (post hoc test showed that set size had no effect on guess rate in 5,000-ms condition), $F(1, 17) = 0.008$, $p = .93$, $\eta_p^2 = .0005$. This suggests that nearly all the items in the sample array could be stored in VWM when a sufficiently long viewing duration was available, and errors might be from complete lapses that were also present for the Set Size 1 condition.

Swap error decreased, but was not eliminated, with longer viewing durations. Note that the swap error rate was always zero in the Set Size 1 condition because there was no chance of this error to occur in that case. There were main effects of both duration, $F(3, 51) = 4.31$, $p = .0090$, $\eta_p^2 = .20$, and set size, $F(1, 17) = 70.8$, $p < .001$, $\eta_p^2 = .81$, on the swap rate, and the interaction was also significant, $F(3, 51) = 4.31$, $p = .0090$, $\eta_p^2 = .20$. While the guess rate reached nearly zero when the viewing duration was 5,000 ms, some proportion of the trials (about 11%) still suffered from a swap error. Swap rate of about 10% to 20% is typically observed in studies using static stimuli such as color and orientation (Bays et al., 2009; Bays et al., 2011). The observation that the swap error was not eliminated even for the long (5 seconds) exposure duration suggests a general limitation of VWM capacity regardless of stimulus type. Or it may be that the object-location binding of BM was relatively difficult, since there was a similar observation
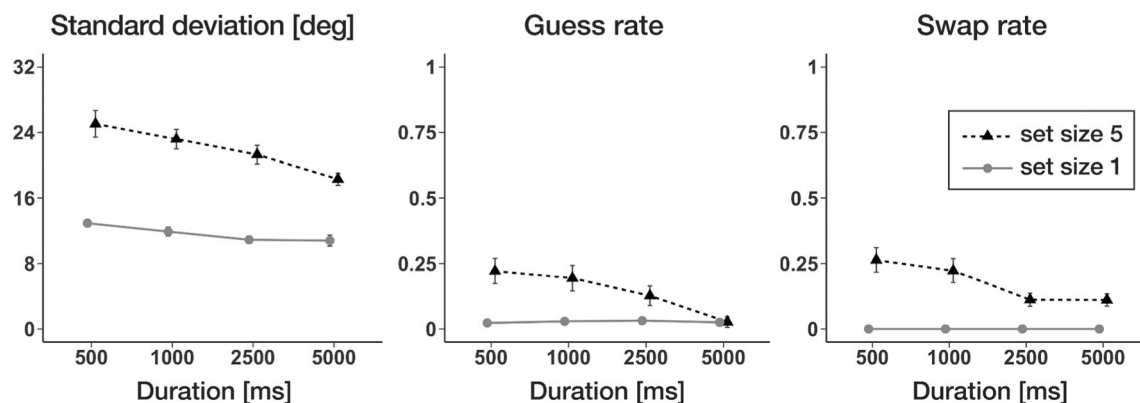
Fig. 2 Results of Experiment 1a. Effects of viewing duration and set size on memory precision, guess rate, and swap rate are shown. Error bars indicate the *SEM*

that the color-action binding of BM was severely limited in capacity (Ding et al., 2015).

Taken together, Experiment 1a demonstrated that when presented with multiple BM items, relatively long periods of presentation time were required to accumulate the precision and capacity of stored items in VWM.

Experiment 1b

The improvements in memory precision with longer viewing durations observed in Experiment 1a may reflect a higher demand on VWM encoding and consolidation processes for complex motion stimuli. However, the duration of several seconds seems to be too long compared with an ordinary time scale for VWM encoding and consolidation. Instead, the performance gain may more likely reflect the limits of perception for the following reasons. First, there was a lack of gait information available when the viewing duration was short: The walker had one gait cycle per second, so participants could observe only half the gait cycle in the 500-ms condition. Second, repeated sampling of gait cycles beyond 1 second will further stabilize and increase the reliability of the motion direction estimate. Therefore, the improvements in recall precision in Experiment 1a may be explained by the temporal limits of the perceptual sensitivity to BM rather than the mnemonic processes such as encoding and consolidation in VWM.

To test this possibility, we conducted a psychophysical experiment where perceptual precision, rather than memory precision, was measured while stimuli and viewing conditions were matched to Experiment 1a. Observers were presented with BM stimuli and they had to discriminate the heading direction thereof. The perceptual sensitivity was evaluated by the difference limen (DL) that was obtained from psychometric curves, and the effects of viewing duration and set size on DL were examined.

Method

Participants, stimuli, and apparatus Eight Kyoto University students (18–22 years old) participated in the experiment for monetary compensation. All participants reported normal color vision and had normal or corrected-to-normal visual acuity. Participants provided written consent to the procedure of the experiment. The sample size was determined on the basis of the effect sizes in Experiment 1a to achieve 80% power by using G*Power (Version 3.1.9.3; Faul, Erdfelder, Lang, & Buchner, 2007). The same stimulus, apparatus, and viewing conditions were used as in Experiment 1a.

Design, procedure, and analysis The experiment used the method of constant stimuli with nine levels of difference

between the walker directions. The standard stimulus was presented with a heading direction of 45° (heading direction was defined as 0° when the walker was facing directly toward the observer and ±90° when it was facing rightward or leftward). The direction of the test stimulus was 45+0° (identical to the standard stimulus), 45±4°, 45±8°, 45±12°, or 45±30° (symmetric design). When the set size was five, the directions of the other walkers (walkers that were not the standard or the test stimulus) were randomly assigned, but with the constraint that they did not affect the correct response defined by the directions of the standard and the test stimuli.

See Fig. 3 for the procedure. Participants were first presented with a central cross for 500 ms against a dark gray background. Then, the standard and test stimuli were presented with varying durations (500; 1,000; 2,500; or 5,000 ms), which were randomly assigned across trials. In the Set Size 2 condition, a standard and a test stimulus was presented side by side (the center of each item was separated by 6.5°), followed by masking noise for 200 ms. Participants were asked to decide which walker was facing more directly toward them (i.e., which walker was close to the direction of 0°). The response was made by a key press. The next trial began 500 ms after the response, and no feedback about the accuracy was given.

In the Set Size 5 condition, a standard stimulus and the other four stimuli (including a test stimulus) were presented with a circular configuration as in Experiment 1a (see Fig.3 for an example display), but the positions of the stimuli were fixed (no shift or jitter of stimulus positions). There were two types of stimulus configurations. The first was as in Fig. 3, where the standard stimulus was positioned at the left-most location of the configuration, and in the second, the configuration was left-right reversed so that the standard stimulus was positioned at the right-most location of the configuration. The stimulus configuration was fixed throughout the task, and each participant was assigned to either one of the configurations. The task was to respond to which group (left or right) a walker was most directly facing toward the observer. When the stimulus configuration was as in Fig. 3 (the standard stimulus was positioned at the left-most location), the left group refers to a walker at the left-most location (i.e., the standard stimulus), and the right group refers to the other four walkers. For a trial as shown in Fig. 3, the lower-right walker is most directly facing toward the observer, which belongs to the right group, so the correct response is to press the right key (response for the right group). There were 12 repetitions for each test stimulus direction, viewing duration, and set size, resulting in a total of 864 trials, with the presentation order and stimulus configuration randomized across participants.

The 25th (P25) and 75th (P75) percentiles of the response distribution were calculated by the Spearman–Kärber method (Miller & Ulrich, 2001). Half the interquartile range, (P75 − P25) / 2, was used to estimate the difference limen (DL), which is inversely related to perceptual sensitivity.

Sample array
(500–5000 ms)

Mask
(200 ms)

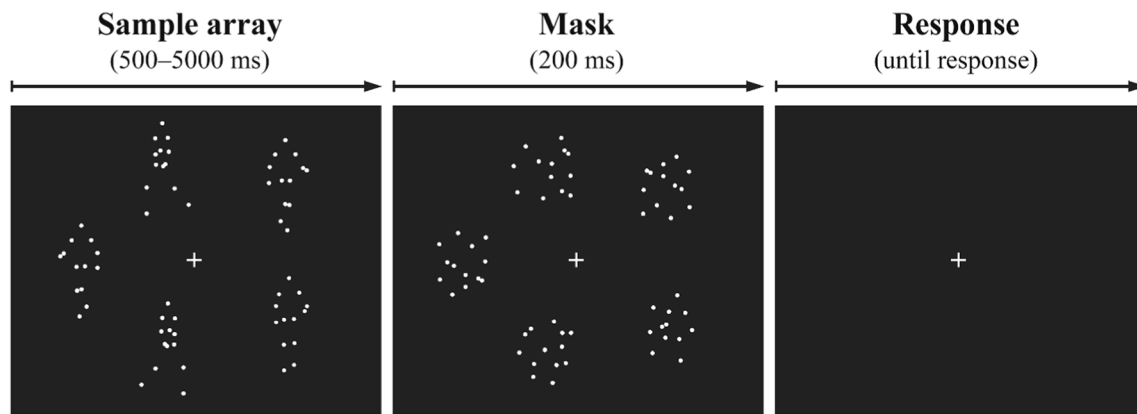## Response
(until response)

Fig. 3 Schematic illustration of a single trial in Experiment 1b

Results and discussion

Psychometric functions and the effects of viewing duration and set size on DL are summarized in Fig. 4. As seen in the figure, the psychometric curves became steeper with increasing viewing duration, indicating a higher perceptual sensitivity at longer viewing durations, which is also reflected in the decrease in DL. There were main effects of both duration, $F(3, 21) = 15.7$, $p < .0001$, $\eta_p^2 = .69$, and set size, $F(1, 7) = 20.2$, $p = .0030$, $\eta_p^2 = .74$, on DL, and the interaction was also significant, $F(3, 21) = 3.92$, $p = .023$, $\eta_p^2 = .36$.

The results showed that the perceptual sensitivity was indeed affected by the viewing duration, and the impact was greater when the set size was large. This temporal profile of perceptual precision was analogous to the results of memory precision observed in Experiment 1a, suggesting that the limits of perception served as a major processing bottleneck in forming detailed VWM representations of the walking directions of BM.

Experiment 2

As shown in Experiment 1, both the precision and the number of stored representations were impaired for shorter viewing durations, and relatively long periods of time were required to fully store multiple motion information in VWM. In Experiment 2, we conducted a series of experiments to examine whether this is specific to BM stimuli or rather a general characteristic of motion stimuli.

The task was almost identical to Experiment 1a, but with the use of different stimulus sets (oriented bar in Experiment 2a, an RDK in Experiment 2b, and BM in Experiment 2c). The bar experiment (Experiment 2a) served as a control experiment in order to compare results between static and dynamic stimuli. The effect of stimulus complexity can be examined by comparing results of the RDK and BM. In these three experiments, a Set Size 3 condition was added to examine the effect of memory load more carefully, and the number of duration conditions was reduced to keep the number of task trials nearly constant.
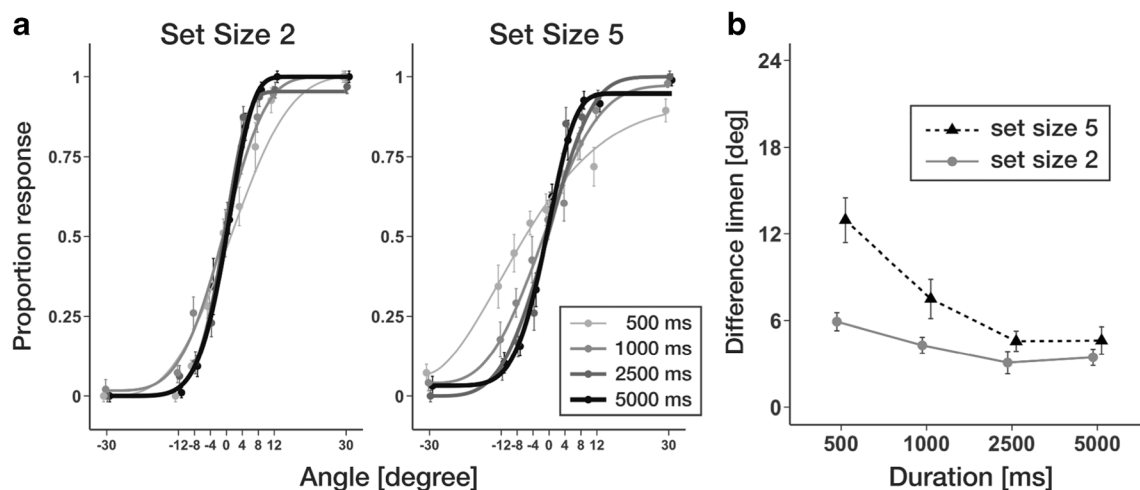

Fig. 4 **a** Psychometric function and (**b**) difference limen (DL) for the discrimination of the heading direction of BM with varying viewing durations. Error bars indicate the *SEM*

Method

Participants Forty-two Kyoto University students (18–22 years old) participated in the experiment for monetary compensation. Participants were assigned to one of the three experiments ($n = 14$ per group). All participants reported normal color vision and had normal or corrected-to-normal visual acuity. Participants provided written consent to the procedure of the experiment. The sample size was determined on the basis of the effect sizes in Experiment 1a to achieve 80% power to detect the effects of viewing duration or set size on each parameter.

Stimuli and apparatus In Experiment 2a, the stimulus was an oriented white bar ($4° \times 0.12°$). In Experiment 2b, the stimulus was an RDK consisting of 30 white dots (each with $0.15°$ diameter) with 100% coherent motion direction (with a constant speed of $4.5°$/s). The dots were displayed within an invisible aperture of $4°$ in diameter, and dots reaching the edge of the aperture were randomly repositioned on the other side of the aperture, keeping the dot density constant throughout the presentation. In Experiment 2c, the stimuli were the same as in Experiment 1a (BM). In all the experiments, the orientations or directions of each stimulus in the sample array were randomly assigned in each trial. The apparatus and viewing conditions were the same as in Experiment 1a.

Design, procedure, and analysis The design and procedure of the task was almost the same as in Experiment 1a, except for the following: (1) the viewing duration was 500; 2,500; or 5,000 ms (the 1,000-ms condition was excluded), and (2) the set size was one, three, or five (a set size of three was included). Participants were asked to remember the orientations of bars (Experiment 2a), motion directions of an RDK (Experiment 2b), or heading directions of BM (Experiment 2c). In the test phase, participants adjusted the orientation or the direction of the probe stimulus so that it matches to that of an item that had been presented at the same location in the memory array.

Each experiment consisted of 459 trials in total (51 trials for each combination of the experimental manipulations). The presentation duration varied between trials and the different set sizes were tested in separate blocks of trials, as in Experiment 1a. There were nine blocks in total, and participants were allowed to rest between the blocks. The analysis was conducted in the same way as in Experiment 1a.

Results and discussion

The results of Experiment 2 are summarized in Fig. 5. We conducted an ANOVA using the same within-subjects factors as in Experiment 1a (viewing duration and set size) with an additional between-subjects factor (stimulus: bar, RDK, or BM).

For the *SD*, there were main effects of stimulus, $F(2, 39) = 33.29$, $p < .0001$, $\eta_p^2 = .63$, duration, $F(2, 78) = 6.19$, $p = .0032$, $\eta_p^2 = .14$, and set size, $F(2, 78) = 305.7$, $p < .0001$, $\eta_p^2 = .89$. While there were significant interactions between stimulus and set size, $F(4, 78) = 9.69$, $p < .0001$, $\eta_p^2 = .33$, and between duration and set size, $F(4, 156) = 3.63$, $p = .0074$, $\eta_p^2 = .085$, the other interactions were not significant, Stimulus × Duration, $F(4, 78) = 1.50$, $p = .21$, $\eta_p^2 = .071$; Stimulus × Duration × Set Size, $F(8, 156) = 1.19$, $p = .31$, $\eta_p^2 = .058$. While the effect of duration was not significant in the case of bar, $F(2, 26) = 0.36$, $p = .70$, $\eta_p^2 = .027$, and RDK, $F(2, 26) = 1.55$, $p = .23$, $\eta_p^2 = .11$, viewing duration had effect for BM, $F(2, 26) = 13.01$, $p = .0001$, $\eta_p^2 = .50$. The interaction between duration and set size was also significant for BM, $F(4, 52) = 5.39$, $p = .0010$, $\eta_p^2 = .29$, and multiple comparisons showed that the *SD* in the 500-ms condition differed from those in 2,500 and 5,000-ms conditions at Set Size 3, and the *SD* in 2,500-ms condition differed from that in 5,000-ms condition at Set Size 5 (all *p*s < .05, Shaffer's Bonferroni corrected).

For guess rate, while there were main effects of duration, $F(2, 78) = 18.3$, $p < .0001$, $\eta_p^2 = .32$, and set size, $F(2, 78) = 83.4$, $p < .0001$, $\eta_p^2 = .68$, the effect of stimulus was not significant, $F(2, 39) = 3.18$, $p = .053$, $\eta_p^2 = .14$. Interactions were all significant, Stimulus × Duration, $F(4, 78) = 6.29$, $p = .0002$, $\eta_p^2 = .24$; Stimulus × Set Size, $F(4, 78) = 10.3$, $p < .0001$, $\eta_p^2 = .33$; Duration × Set Size, $F(4, 156) = 18.8$, $p < .0001$, $\eta_p^2 = .33$; Stimulus × Duration × Set Size, $F(8, 156) = 5.25$, $p < .0001$, $\eta_p^2 = .21$. While the effect of duration was not significant in the case of bar, $F(2, 26) = 0.84$, $p = .44$, $\eta_p^2 = .061$, viewing duration had effect for RDK, $F(2, 26) = 23.4$, $p < .0001$, $\eta_p^2 = .64$, and for BM, $F(2, 26) = 18.2$, $p < .0001$, $\eta_p^2 = .58$, and the interaction between duration and set size was also significant for RDK, $F(4, 52) = 16.1$, $p < .0001$, $\eta_p^2 = .55$, and for BM, $F(4, 52) = 8.91$, $p < .0001$, $\eta_p^2 = .41$. Multiple comparisons showed that guess rate in the 500-ms condition differed from those in 2,500 and 5,000-ms conditions at Set Size 5, in the case of both RDK and BM (all *p*s < .05, Shaffer's Bonferroni corrected).

For swap rate, there were main effects of stimulus, $F(2, 39) = 3.51$, $p = .040$, $\eta_p^2 = .15$; duration, $F(2, 78) = 4.10$, $p = .020$, $\eta_p^2 = .095$; and set size, $F(2, 78) = 70.5$, $p < .0001$, $\eta_p^2 = .64$. Interactions were all significant, Stimulus × Duration, $F(4, 78) = 4.94$, $p = .0013$, $\eta_p^2 = .20$; Stimulus × Set Size, $F(4, 78) = 3.25$, $p = .016$, $\eta_p^2 = .14$; Duration × Set Size, $F(4, 156) = 2.52$, $p = .043$, $\eta_p^2 = .061$; Stimulus × Duration × Set Size, $F(8, 156) = 2.67$, $p = .0088$, $\eta_p^2 = .12$. While the effect of duration on guess rate was not significant in the case of bar, $F(2, 26) = 1.52$, $p = .23$, $\eta_p^2 = .11$, and RDK, $F(2, 26) = 0.32$, $p = .72$, $\eta_p^2 = .025$, viewing duration had effect for BM, $F(2, 26) = 7.18$, $p = .0033$, $\eta_p^2 = .36$. The interaction between duration and set size was also significant for BM, $F(4, 52) = 4.38$, $p = .0040$, $\eta_p^2 = .25$, and multiple comparisons showed that the *SD* in the 500-ms condition differed from that in the
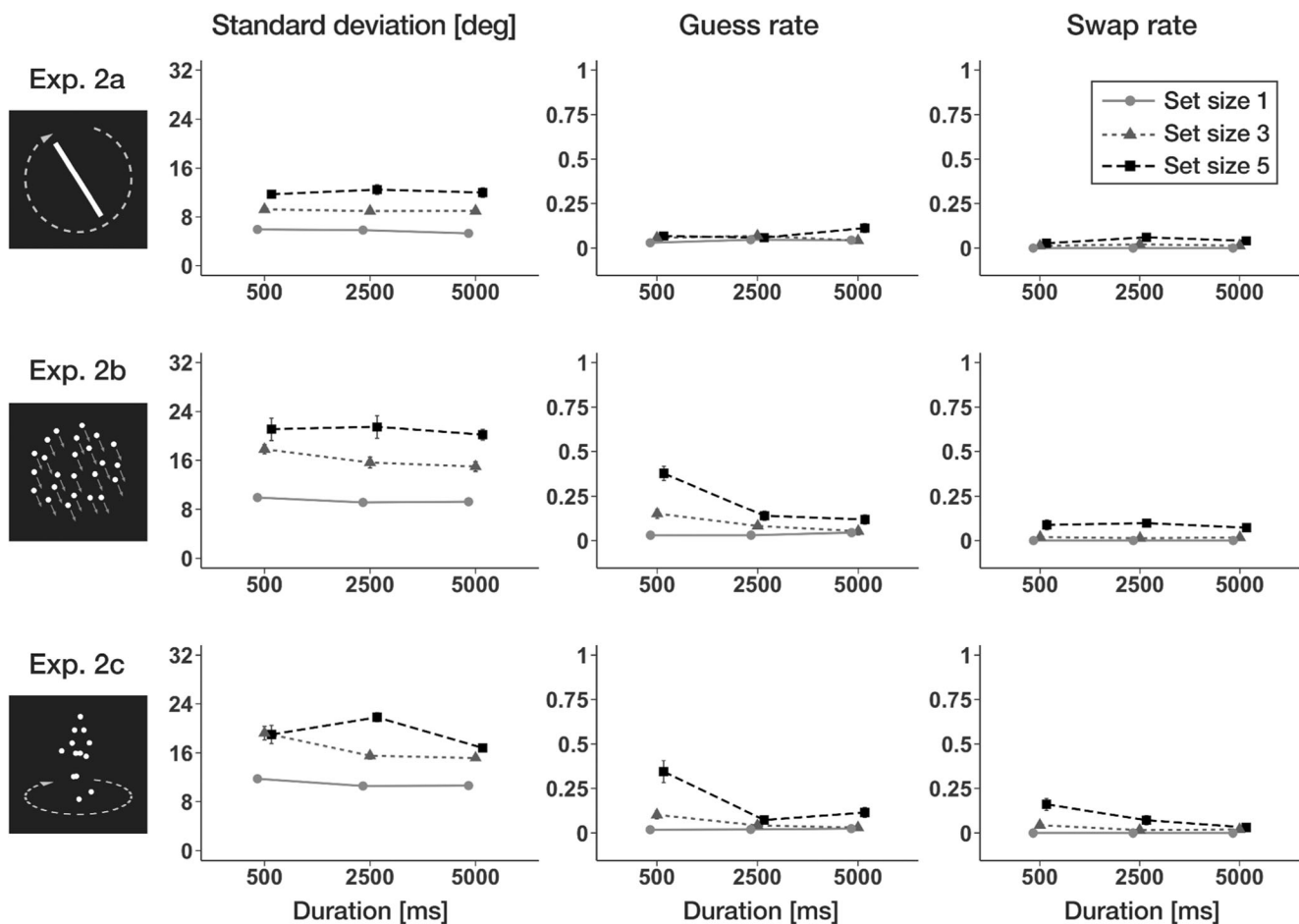
Fig. 5 Experiment 2. Effects of viewing duration and set size on memory precision, guess rate, and swap rate for different types of stimuli are shown: (top row) oriented bars (Experiment 2a), (middle row) random dot kinematograms (RDKs; Experiment 2b), and (bottom row) biological motion (BM; Experiment 2c). Error bars indicate the *SEM*

5,000-ms condition at Set Size 5 (p = .010, Shaffer's Bonferroni corrected). In the case of RDK, the effect of set size was significant, $F(2, 26) = 26.33$, $p < .0001$, $\eta_p^2 = .64$.

In summary, while memory of bar orientation did not benefit from prolonged viewing, memory of complex motion (BM) gradually improved as the viewing duration increased in terms of both precision and capacity measures, and the result of RDK was in between: Only the number of stored items was affected by viewing duration. For both types of motion stimuli (RDK and BM), the number of stored items was considerably decreased in the 500-ms condition when the set size was five. A viewing duration of at least 2,500 ms was required for guess rate to reach its asymptotic performance. In contrast, the successful storage of five static orientations was well achieved within 500 ms of presentation duration. These results demonstrate that it requires relatively long periods of time, regardless of stimulus complexity, to store multiple motion directions in VWM.

Experiment 3a

As observed in Experiment 2, the development of memory precision was slow only for BM. We then asked which critical stimulus property accounts for the difference in information processing speed between BM and an RDK. We assumed that the critical difference between BM and the RDK was their visual complexity: BM is a structured stimulus that requires the integration of local motion signals to perceive a globally coherent motion pattern, which also yields structural (shape) information (Blake & Shiffrar, 2007; Chang & Troje, 2009). However, there might be an alternative explanation for the difference in results between BM and the RDK. When perceiving multiple motion items, it is known that the processing capacity is greatly affected by the signal-to-noise level of the stimulus—the increase in stimulus noise decreases the number of localized motion signals that can be simultaneously perceived (Edwards & Rideaux, 2013; Greenwood & Edwards, 2009). Since the BM stimuli consist of dots with a distinct spatiotemporal energy, higher internal noise is produced in

the motion system, which can induce a low ceiling on motion processing efficiency (Watamaniuk, 1993). Therefore, rather than the complex structured nature of the stimulus itself, noisy input to the motion system alone may be sufficient to cause the inefficient accumulation of memory precision in VWM. If this is the case, encoding of the RDK would be also inefficient when noise is introduced into the stimulus.

To investigate this possibility, we manipulated the signal-to-noise levels, or the motion coherence of the RDK. We used an RDK with 80% coherence (20% of dots that constitute an RDK were assigned random directions) as a noisy motion stimulus and compared the results with a noiseless counterpart (i.e., RDK with 100% coherence, which is what we used in Experiment 2b). If the signal-to-noise level of the stimulus is a crucial factor for determining the processing efficiency of motion information in VWM, the performance will be impaired by the presence of noise. The experimental procedure was almost identical to Experiment 1a, except that the set size was fixed to five.

Method

Participants, stimuli, and apparatus Sixteen Kyoto University students (18–22 years old) participated in the experiment for monetary compensation. All participants reported normal color vision and had normal or corrected-to-normal visual acuity. Participants provided written consent to the procedure of the experiment. The noiseless RDK had 100% coherence, and the noisy RDK had 80% coherence (20% of the dots were assigned random directions). The apparatus and viewing conditions were identical to those in Experiment 1a.

Design, procedure, and analysis The design and procedure of the task were almost identical to those in Experiment 1a, except for the following: (1) An RDK was used as the stimulus, and (2) set size was fixed to five. The presentation duration varied between trials, and the different coherences were tested in separate blocks of trials. The probe stimulus was the noiseless RDK regardless of whether the preceding samples were noisy or not.

Results and discussion

The results are summarized in Fig. 6. The signal-to-noise levels had an effect on the guess rate, but not on the SD or the swap rate. There were main effects of both the coherence, $F(1, 15) = 12.5, p = .0030, \eta_p^2 = .45$, and the duration, $F(3, 45) = 11.3, p < .0001, \eta_p^2 = .43$, on the guess rate, and there was no interaction, $F(3, 45) = 0.35, p = .79, \eta_p^2 = .023$. For the results of the SD and the swap rate, there was no effect of the coherence, $F(1, 15) = 0.01, p = .92, \eta_p^2 = .0007$, or the viewing duration, $F(3, 45) = 1.00, p = .40, \eta_p^2 = .062$, on the SD, and there was no interaction, $F(3, 45) = 0.067, p = .98, \eta_p^2 = $

.0045. Similarly, there was no effect of the coherence, $F(1, 15) = 0.35, p = .56, \eta_p^2 = .023$, or the viewing duration, $F(3, 45) = 0.27, p = .84, \eta_p^2 = .018$, on the swap rate, and there was no interaction, $F(3, 45) = 0.27, p = .84, \eta_p^2 = .018$.

As shown in the results section, the presence of noise indeed impaired the memory performance. However, the effect of stimulus noise was reflected in the number, rather than the precision, of stored items in VWM. There was an overall decline in the guess rate for the noisy RDK.

Experiment 3b

As in Experiment 1, we also examined the perceptual sensitivity to the RDK and how it was affected by the viewing duration and the signal-to-noise levels in order to better understand the results of Experiment 3a.

Method

Participants, stimuli, and apparatus Eight Kyoto University students (18–22 years old) participated in the experiment for monetary compensation. All participants reported normal color vision and had normal or corrected-to-normal visual acuity. Participants provided written consent to the procedure of the experiment. The stimuli, apparatus, and viewing conditions were identical to those of Experiment 3a.

Design, procedure, and analysis The task was almost identical to that of Experiment 1b. Instead of the set size, in this case the motion coherence of the RDK was manipulated. The set size was fixed to five, and the participants were asked to respond in which group (standard or the other four stimuli) the RDK had the most downward (6 o'clock) direction. There were 12 repetitions for each combination of direction (0°, ±4°, ±8°, ±12°, and ±30°), duration (500; 1,000; 2,500; and 5,000 ms), and coherence (100% and 80%), resulting in a total of 864 trials. The data analysis was conducted the same way as in Experiment 1b.

Results and discussion

Psychometric functions and the effects of viewing duration and motion coherence on DL are summarized in Fig. 7. There was no effect of the coherence on DL, $F(1, 15) = 0.0021, p = .97, \eta_p^2 = .0003$. There was a main effect of duration, $F(3, 45) = 6.31, p = .0032, \eta_p^2 = .47$, and no interaction was present, $F(3, 45) = 1.09, p = .38, \eta_p^2 = .13$. In comparison with BM (Experiment 1b), the processing of the RDK was relatively fast. In addition, the perceptual sensitivity to the RDK was not affected by the presence of stimulus noise. The results of Experiments 3a and 3b suggest that because the presence of noise influenced memory performance in a
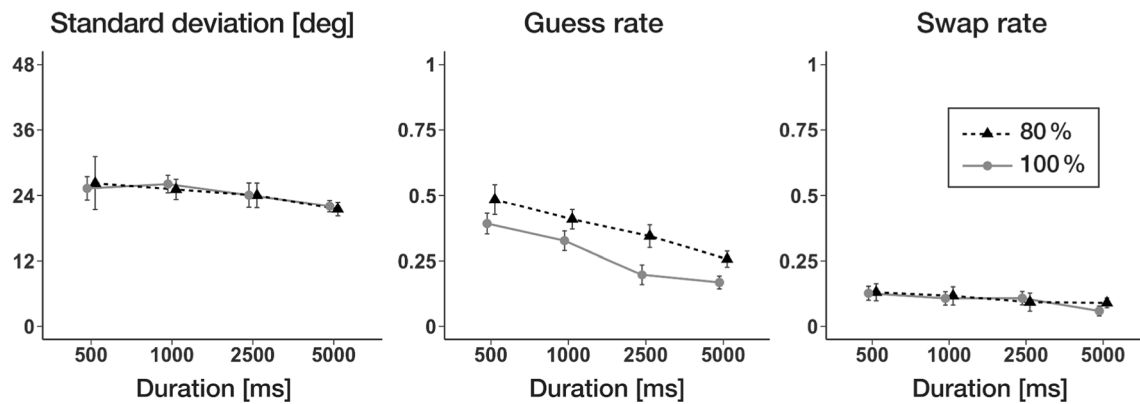
Fig. 6 Experiment 3a. Effects of viewing duration and motion coherence on memory precision, guess rate, and swap rate are shown. Error bars indicate the *SEM*

qualitatively different fashion, the inefficiency of forming VWM for BM stimuli cannot be explained by the decreased signal-to-noise level of the stimulus.

General discussion

In the present study, we investigated the temporal development of VWM representations of motion direction in terms of both the precision and the number of stored items. Across experiments, the gradual formation of VWM representations of motion direction was observed. The memory performance continued to improve with the increase in viewing duration over relatively long periods of time of up to several seconds. For BM stimuli, the improvements were reflected in both precision and capacity measures (Experiment 1), while for the RDK, only the capacity (the number of stored items in VWM) was affected by viewing duration (Experiments 2 and 3). We also found that the decrease in the signal-to-noise levels of the RDK led to the decrease in the number of stored items, but not in memory precision (Experiment 3). By examining the temporal profiles of perceptual sensitivity to the

stimuli, we found that while it required relatively long periods of time to acquire a precise estimate of heading direction of BM, the perception of the RDK was relatively quick, regardless of the noise levels. Taken together, these results characterized limitations underlying the formation of VWM representations for simultaneously presented dynamic objects.

Why is the formation of VWM representations of motion information so much slower than that of simple and static features such as line orientation, which can be stored within several hundred milliseconds (Experiment 2a; see also Bays et al., 2011, for a similar observation)? To form a fine-grained representation of motion direction, each object must be accurately and precisely perceived, encoded, and consolidated in VWM, and all these processes seem to be less efficient for motion stimuli. First, perceptual speed for motion direction was relatively slow. The perception of BM and the RDK required at least 1 to 2 seconds of viewing duration to reach asymptotic sensitivity (Experiments 1b and 3b). Perceptual performance was also worse for larger set sizes, suggesting an increased difficulty of simultaneous processing of multiple motion stimuli. As a result, in Experiment 2, we observed that storing of multiple motion directions was especially challenging
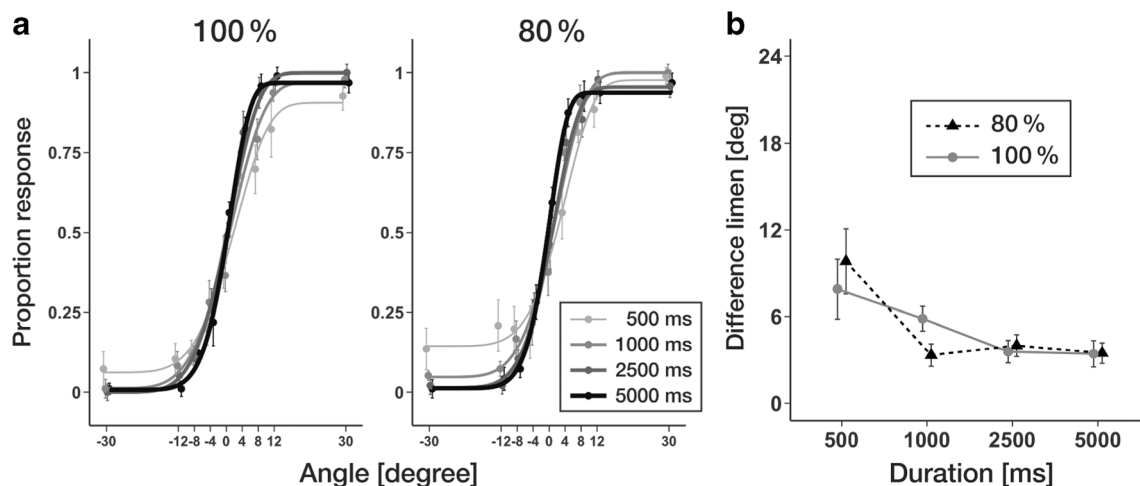
Fig. 7 **a** Psychometric function and (**b**) difference limen (DL) for discriminating motion direction of the RDK. Error bars indicate the *SEM*

when the viewing duration was short (500 ms) for both types of motion stimuli (BM and RDK), as reflected in high guess rates. The accumulation of memory precision was also gradual for BM stimuli. Although a binary discrimination of walking direction (left or right) of a point-light walker can be achieved for exposure times as short as 200 ms (Chang & Troje, 2008, 2009), much longer time was required to achieve a detailed perception of the walking direction of BM (Experiment 1b), which would explain the remarkably gradual formation of the fine-detailed memory of the walking direction of BM.

Second, as suggested by Rideaux et al. (2015), encoding and consolidation of motion information in VWM also appear to be slow. Although the difference in the processing time between static and motion stimuli may not be so large (in the range of several dozen milliseconds; Rideaux et al., 2015), the inefficiencies in encoding and consolidation will increasingly become salient as the number of displayed items increases. Recent studies have shown that parallel consolidation in VWM is severely limited in capacity, with only one or two items that can be processed at a time, regardless of whether the stimulus was a static orientation or a motion direction (Becker et al., 2013; Mance et al., 2012; Rideaux et al., 2015). Therefore, the delay in total processing time to complete the storage of whole items will increase for larger stimulus set sizes in the display, due to the repetitions of (inefficient) encoding and consolidation cycles for motion stimuli.

As a result, the inefficiencies in the construction of VWM representations of motion information reflect the limits of several processes from perception to memory, rather than a single property of a specific processing stage such as consolidation. Another potential explanation that might be raised for the performance gains for longer viewing durations is a contribution from elaborative encoding, such as grouping similar items into a chunk. A viewing duration of several seconds can be long enough to introduce such strategic effects. If so, performance should be facilitated for longer viewing conditions regardless of the stimulus type. However, the viewing duration did not affect memory performance for line orientation (see Experiment 2a, although the performance could have reflected a ceiling effect and therefore strategic modulation could have little effect on that task). In addition, since the viewing duration was randomly assigned for each trial, the participants could not predict how long the stimuli would remain in the display. Therefore, it was not possible to switch the encoding strategy depending on the presentation duration of the sample array. Considering the temporal limits for perceiving multiple motion items as observed in the current study with psychophysical tasks, along with the slower consolidation rate for motion stimuli (Rideaux et al., 2015), we argue that inefficiencies in the information flow at both perceptual and memory formation stages were a major factor for the gradual formation of VWM representations of dynamic stimuli.

The memory capacity was reduced by increasing the noise levels of the RDK (Experiment 3a). It has been previously reported that the signal-to-noise level affects the number of motion directions that can be simultaneously perceived (Edwards & Rideaux, 2013; Greenwood & Edwards, 2009). In the current study, however, the perceptual sensitivity to the motion direction of the RDK was not affected by noise levels (Experiment 3b). Therefore, the decrease in memory capacity may reflect the impaired performance of postperceptual stages such as encoding/consolidation or maintenance. The signal-to-noise levels did not affect the perceptual sensitivity to the RDK, presumably because the degree of noise level was relatively small. In fact, the motion coherence of 80% is much higher than that used in typical motion perception studies, and the threshold of detecting the motion direction of the RDK can be as small as 5% (Britten, Shadlen, Newsome, & Movshon, 1992). If we had used stimuli with lower motion coherence, the memory performance could have likely been impaired not only in capacity but in precision as well. For the RDK, we used a motion coherence of 80% based on a pilot study that showed that even an RDK with 60% coherence was too noisy to perform the task of our current study confidently. We believe that the lower tolerance to stimulus noise in the current experiment may be due to higher task demands. In contrast to typical perceptual tasks that require only a binary discrimination of motion direction (leftward or rightward) of one or two RDKs, the detailed discrimination of multiple RDKs and their maintenance in memory in our task was more demanding for the participants. Therefore, we do not intend to argue that the signal-to-noise levels will exclusively affect the guess rate, since memory precision can also be impaired depending on noise levels. Instead, what we want to highlight from the present results is that the signal-to-noise level was more sensitive to the capacity measure than to the precision (i.e., even a small noise that did not affect perceptual sensitivity could nevertheless impair memory capacity). This observation indicates qualitatively distinct effects of stimulus noise and visual complexity on VWM performance, suggesting that the gradual formation of VWM representations for BM stimuli, as compared with those for an RDK, cannot be explained by the presence of noise alone. The underlying mechanisms for the effects of the signal-to-noise level on memory processes and the interaction with stimulus complexity still remain largely unclear, and future work is needed to address these issues. In addition, but related, although the analysis of the current study was based on the mixture model of VWM, a recent paper by Schurgin, Wixted, and Brady (2018) disputes the model and demonstrated that once perceptual space has been accommodated, "guessing" need not be incorporated to model the result of a continuous report task, arguing against a view that precision and capacity are independent measures that reflect different aspects of memory state. The idea of incorporating perceptual space into account would be a valuable approach too

in designing future research that examines the effect stimulus type on VWM.

Our observations may provide an explanation for an apparent discrepancy about the capacity limit of VWM for dynamic objects. For example, while Kawasaki, Watanabe, and Okuda (2008) reported that participants could retain only two RDKs in VWM, Shen et al. (2014) showed that three to four BMs could be retained in VWM. In the former study, four items were presented simultaneously for 250 or 500 ms, whereas in the latter study the viewing duration increased by N seconds when the set size was N (1 s/item). In Experiment 2 of our study, with a set size of five, a viewing duration of 500 ms was insufficient to fill the capacity of VWM for both the RDK and BM. Therefore, a 250–500-ms presentation duration in Kawasaki et al. (2008) may be too short for the adequate encoding of simultaneously presented dynamic stimuli. Our study supports the view that the capacity of VWM for dynamic objects is comparable with that for simple static features (Blake, Cepeda, & Hiris, 1997; Shen et al., 2014; Wood, 2007) and emphasizes the necessity of minimizing encoding limitations when estimating VWM capacity for dynamic stimuli.

VWM tends to be poorer for complex relative to simple items (Alvarez & Cavanagh, 2004), especially when stimulus presentation duration is short (Eng et al., 2005) or sample-test similarity is high (Awh, Barton, & Vogel, 2007; Jackson, Linden, Roberts, Kriegeskorte, & Haenschel, 2015). However, by contrast to lower storage capacity that is observed with complex but meaningless items (i.e., random polygons or shaded cubes), capacity of visual memory for complex but meaningful objects such as faces and real-world objects is comparable with or even larger than that observed for simple items (Brady, Störmer, & Alvarez, 2016; Curby & Gauthier, 2007; Endress & Potter, 2014). Additional encoding time does not improve memory for simple stimuli (Brady et al., 2016; Luck & Vogel, 1997) but it does for complex stimuli, especially when they are meaningful for observers (Curby, Glazek, & Gauthier, 2009). Several factors seem to be relevant for the benefit of prolonged viewing for complex but meaningful objects: the involvement of long-term memory (Makovski & Jiang, 2008), contributions from semantic or conceptual knowledge (Chase & Simon, 1973; Ericsson & Kintsch, 1995), and the efficient encoding of objects within one's domain of expertise (Curby et al., 2009). Of these factors, perhaps the most relevant to the current work is the effect of perceptual expertise on VWM (Curby et al., 2009). This view argues that there is a VWM advantage for objects within one's domain of expertise (e.g., faces) that is driven by specialized processing strategy for those objects (holistic processing). This advantage requires sufficient encoding time (Curby & Gauthier, 2007; Curby et al., 2009). Much like for faces, human vision is expertized for human motion processing (Blake & Shiffrar, 2007). As

suggested for faces, the gradual improvement of VWM for BM might be supported by the efficient encoding mechanisms, which leads to the formation of accurate representation, but with sufficient encoding time. By contrast to BM, RDK has no meaningful structure. Once perceptual limitation is eliminated, additional encoding time may have no functional role in improving the quality of representation of RDK.

In conclusion, our current study demonstrated some remarkable temporal limitations in the formation of VWM representations of motion information, especially when the stimulus is visually complex. Long viewing durations are required to accurately retain information about dynamic objects. The immediate availability of information in the outside world may in most cases hold true for static stimuli, but this assumption is challenged by dynamic scenes, especially when the precise estimate of the feature value is important. Our results have implications for understanding how people interact with dynamic environments. For example, when we are walking along in a crowd, active monitoring and early identification of pedestrians in the surrounding area is important to avoid potential collisions (Foulsham, Walker, & Kingstone, 2011; Jovancevic, Sullivan, & Hayhoe, 2006). Even though the human visual system is efficient for perceiving the gist of a crowded environment with objects and agents (Sweeny, Haroz, & Whitney, 2013; Whitney & Yamanashi Leib, 2018), we often bump into someone on the street or at the station. This could be due to the difficulty in encoding and maintenance of multiple walker information with sufficient precision for collision avoidance.

References

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. Psychological Science, 15(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. Psychological Science, 18(7), 622–628. https://doi.org/10.1111/j.1467-9280.2007.01949.x

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory representations in natural tasks. Journal of Cognitive Neuroscience, 7(1), 66–80. https://doi.org/10.1162/jocn.1995.7.1.66

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. Journal of Vision, 9(10):7.1–11, https://doi.org/10.1167/9.10.7

Bays, P. M., Gorgoraptis, N., Wee, N., Marshall, L., & Husain, M. (2011). Temporal dynamics of encoding, storage, and reallocation of visual

working memory. *Journal of Vision, 11*(10):1–15. https://doi.org/10.1167/11.10.6

Becker, M. W., Miller, J. R., & Liu, T. (2013). A severe capacity limit in the consolidation of orientation information into visual short-term memory. *Attention, Perception, & Psychophysics, 75*(3), 415–425. https://doi.org/10.3758/s13414-012-0410-0

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motion. *Psychological Science, 5*, 221–225. https://doi.org/10.1111/j.1467-9280.1994.tb00504.x

Blake, R., Cepeda, N. J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology: Human Perception and Performance, 23*(2), 353–369. https://doi.org/10.1037/0096-1523.23.2.353

Blake, R., Shiffrar, M. (2007). Perception of human motion. *Annual Review of Psychology, 58*:47-73. https://doi.org/10.1146/annurev.psych.57.102904.190152

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: More active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences, 113*(27), 7459–7464. https://doi.org/10.1073/pnas.1520027113

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual motion: A comparison of neuronal and psychophysical performance. *The Journal of Neuroscience, 12*(12), 4745–4765. https://doi.org/10.1523/jneurosci.12-12-04745.1992

Chang, D. H. F., & Troje, N. F. (2008). Perception of animacy and direction from local biological motion signals. *Journal of Vision, 8*(5): 3.1–10. https://doi.org/10.1167/8.5.3

Chang, D. H. F., & Troje, N. F. (2009). Characterizing global and local mechanisms in biological motion perception. *Journal of Vision, 9*(5):8.1–10, https://doi.org/10.1167/9.5.8

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*(1), 55–81. https://doi.org/10.1016/0010-0285(73)90004-2

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic Bulletin & Review, 14*(4), 620–628. https://doi.org/10.3758/bf03196811

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 94–107. https://doi.org/10.1037/0096-1523.35.1.94

Ding, X., Zhao, Y., Wu, F., Lu, X., Gao, Z., & Shen, M. (2015). Binding biological motion and visual features in working memory. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 850–865. https://doi.org/10.1037/xhp0000061

Edwards, M., & Rideaux, R. (2013). How many motion signals can be simultaneously perceived? *Vision Research, 76*, 11–16. https://doi.org/10.1016/j.visres.2012.10.004

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General, 143*(2), 548–565. https://doi.org/10.1037/a0033934

Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex visual stimuli. *Psychonomic Bulletin & Review, 12*(6), 1127–1133. https://doi.org/10.1167/5.8.611

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211–245. https://doi.org/10.1037/0033-295x.102.2.211

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research, 51*(17), 1920–1931. https://doi.org/10.1016/j.visres.2011.07.002

Gao, T., Gao, Z., Li, J., Sun, Z., & Shen, M. (2011). The perceptual root of object-based storage: An interactive model of perception and visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 37*, 1803–1823. doi:https://doi.org/10.1037/a0025637

Gao, Z., Ding, X., Yang, T., Liang, J., & Shui, R. (2013). Coarse-to-fine construction for high-resolution representation in visual working memory. *PLOS ONE, 8*(2), e57913. https://doi.org/10.1371/journal.pone.0057913

Greenwood, J. A., & Edwards, M. (2009). The detection of multiple global directions: Capacity limits with spatially segregated and transparent-motion signals. *Journal of Vision, 9*(1), 40–40. doi: https://doi.org/10.1167/9.1.40

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804. https://doi.org/10.1016/s0896-6273(02)01091-7

Jackson, M. C., Linden, D. E. J., Roberts, M. V., Kriegeskorte, N., & Haenschel, C. (2015). Similarity, not complexity, determines visual working memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(6), 1884–1892. https://doi.org/10.1037/xlm0000125

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201–211. https://doi.org/10.3758/BF03212378

Jolicoeur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology, 36*(2), 138–202. https://doi.org/10.1006/cogp.1998.0684

Jovancevic, J., Sullivan, B., & Hayhoe, M. (2006). Control of attention and gaze in complex environments. *Journal of Vision, 6(12)*, 1431–1450. https://doi.org/10.1167/6.12.9

Kawasaki, M., Watanabe, M., & Okuda, J. (2008). Human posterior parietal cortex maintains color, shape and motion in visual short-term memory. *Brain Research, 13*, 4–6. doi:https://doi.org/10.1016/j.brainres.2008.03.037

Levin, D. T., Momen, N., Drivdahl, S. B., IV, & Simons, D. J. (2000). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition, 7*, 397–412. https://doi.org/10.1080/135062800394865

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281. https://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 3911–400. https://doi.org/10.1016/j.tics.2013.06.006

Makovski, T., & Jiang, Y. V. (2008). Proactive interference from items previously stored in visual working memory. *Memory & Cognition, 36*(1), 43–52. https://doi.org/10.3758/mc.36.1.43

Mance, I., Becker, M. W., & Liu, T. (2012). Parallel consolidation of simple features into visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance, 38*(2), 429–438. doi:https://doi.org/10.1037/a0023925

Miller, J., & Ulrich, R. (2001). On the analysis of psychometric functions: The Spearman-Karber method. *Perception & Psychophysics, 63*, 1399–1420. https://doi.org/10.3758/BF03195545

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology, 46*, 461–88. https://doi.org/10.1037/h0084327

Orhan, A. E., & Jacobs, R. A. (2014). Toward ecologically realistic theories in visual short-term memory research. *Attention, Perception, & Psychophysics, 76*(7), 2158–2170. https://doi.org/10.3758/s13414-014-0649-8

Pinto, Y., Sligte, I. G., Shapiro, K. L., & Lamme, V. A. F. (2013). Fragile visual short-term memory is an object-based and location-specific store. *Psychonomic Bulletin & Review, 20*(4), 732–739. https://doi.org/10.3758/s13423-013-0393-4

Poom, L. (2012). Memory of gender and gait direction from biological motion: Gender fades away but directions stay. *Journal of Experimental Psychology: Human Perception and Performance, 38*(5), 1091–1097. https://doi.org/10.1037/a0028503

Ricker, T. J. (2015). The role of short-term consolidation in memory persistence. *AIMS Neuroscience, 2*(4), 259–279. https://doi.org/10.3934/neuroscience.2015.4.259

Rideaux, R., Apthorp, D., & Edwards, M. (2015). Evidence for parallel consolidation of motion direction and orientation into visual short-term memory. *Journal of Vision, 15*(2), 17. https://doi.org/10.1167/15.2.17

Schurgin, M. W., Wixted, J. T., & Brady, T. F. B. (2018). Psychological scaling reveals a single parameter framework for visual working memory. *BioRxiv.* https://doi.org/10.1101/325472

Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of American Statistical Association, 81*, 826–831. https://doi.org/10.1080/01621459.1986.10478341

Shen, M., Gao, Z., Ding, X., Zhou, B., & Huang, X. (2014). Holding biological motion information in working memory. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1332–1345. https://doi.org/10.1037/a0036839

Sligte, I. G., Scholte, H. S., & Lamme, V. A. F. (2008). Are there multiple visual short-term memory stores? *PLOS ONE, 3*(2), e1699. https://doi.org/10.1371/journal.pone.0001699

Smyth, M. M., Pearson, N. A., & Pendleton, L. R. (1988). Movement and working memory: Patterns and positions in space. The Quarterly Journal of Experimental Psychology Section A, *40*(3), 497–514. https://doi.org/10.1080/02724988843000041

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance, 39*(2), 329–337. https://doi.org/10.1037/a0028712

Thurman, S. M., & Grossman, E. D. (2008). Temporal "bubbles" reveal key features for point-light biological motion perception. *Journal of Vision, 8*(3), 28.1–11. https://doi.org/10.1167/8.3.28

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers, 36*, 625– 629. https://doi.org/10.3758/BF03206542

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1436-1451. https://doi.org/10.1037/0096-1523.32.6.1436

Watamaniuk, S. N. J. (1993). Ideal observer for discrimination of the global direction of dynamic random-dot stimuli. *Journal of the Optical Society of America A, 10*(1), 16. https://doi.org/10.1364/josaa.10.000016

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology, 69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology: General, 136*(4), 639–652. doi:https://doi.org/10.1037/0096-3445.136.4.639

Xu, Y., & Chun, M. M. (2009). Selecting and perceiving multiple visual objects. *Trends in Cognitive Sciences, 13*(4), 167–174. https://doi.org/10.1016/j.tics.2009.01.008

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(May), 233–236. https://doi.org/10.1038/nature06860