Effects of Time Delay on Depth Perception via Head-Motion Parallax in Virtual Environment Systems

Hanfeng Yuan
W. L. Sachtler
Nat Durlach
Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139 USA
Barbara Shinn-Cunningham
Cognitive and Neural Systems
Boston University

Abstract

Experiments were conducted to determine how the ability to detect and discriminate head-motion parallax depth cues is degraded by time delays between head movement and image update. The stimuli consisted of random-dot patterns that were programmed to appear as one cycle of a sinusoidal grating when the subject's head moved. The results show that time delay between head movement and image update has essentially no effect on the ability to discriminate between two such gratings with different depth characteristics when the delay is less than or equal to roughly 265 ms.

1 Introduction

The visual perception of depth—and the resulting perception that the space being viewed is three-dimensional—is often regarded as a significant contributor to the sense of presence in VEs and, in fact, is often included as one of the defining characteristics of VEs.

Although many different types of cues can elicit the perception of depth (including binocular disparity or stereopsis, occlusion of one object by another, geometric perspective, and texture, shading, and size gradients), attention is confined in this paper to depth perception via head-motion parallax. (In the remainder of this paper, we refer to *head-motion parallax* simply as *motion parallax.*)

Depth perception via motion parallax is similar to depth perception via stereopsis in the sense that the perception of depth arises in both cases from the availability of views of a common object or scene from different points in space. In stereopsis, the different viewing points correspond to the locations of the two eyes and the perception of depth arises from interocular comparisons. In motion parallax (which is fundamentally a monocular phenomenon), the different viewing points are generated by relative motion between the object and the eye.

The case of interest to us in this note, illustrated for a single eye, is shown in figure 1a. If the eye is fixated on point F and the eye is moved to the left (maintaining fixation on F), a fixed stimulus at point A, which is farther away than F, appears to move to the left (in the same direction as the eye), whereas a fixed stimulus at point B, which is nearer than F, appears to move to the right (in the direction opposite the eye movement). As illustrated in figure 1b, the magnitude or speed of the movement increases with an increase in the distance of the stimulus location from the fixation point. One can easily observe this phenomenon (which is a direct consequence of the underlying geometry) by holding up a finger, fixating on it, moving one's head from side to side, and noting the movements of objects located in front and behind the finger. It is also easy to note that the phenomenon is essentially unaffected by closing one eye.

To achieve motion parallax in a VE, the system must track the location of the head and update the visual scene appropriately. In addition to the degradations that can occur because of simplifications in the required

Presence, *Vol. 9, No. 6, December 2000, 638–647*

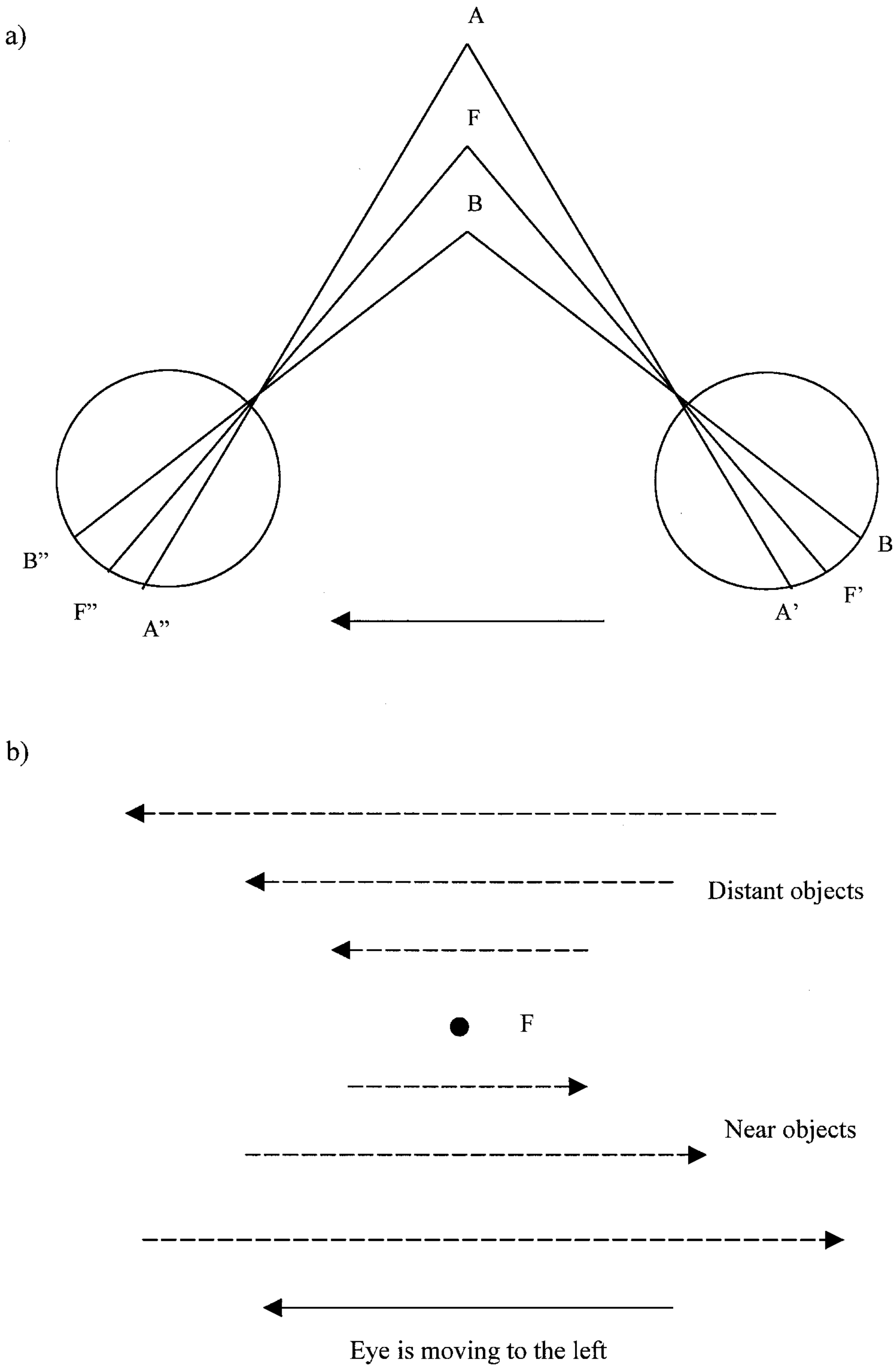

Figure 1. *Schematic illustration of basic geometric factors underlying depth perception by motion parallax. Figure 1a shows movement of retinal images of points A and B when the eye is fixated on an intermediate point F and moves to the left. Relative to point F, point A appears to move to the left and point B to the right. Figure 1b illustrates how the amount of apparent movement of such points increases as the distance between the given point and the fixation point F increases.*

computations or because of imperfections in the visual display, degradations can occur because of time delays between the head movement and the updating of the visual stimulus.

In this note, we report briefly on some experiments that were conducted to examine the effect of time delay on the perception of depth by motion parallax in a desktop VE system. More specifically, the study is concerned with the effect of time delay on the ability to detect and discriminate motion parallax depth cues.

Depth perception via motion parallax has been explored in a variety of previous studies. (See, for example, the results and references in Ferris, 1972; Rogers and Graham, 1979, 1983; Ono and Steinbach, 1990; Hoshino, Miruma, Yamada, and Fukuda, 1991; Satoh, Tomono, and Kishino, 1991; and McCandless, Ellis, and Adelstein, 2000.) Also, the effects of various kinds of time delays on various types of user performance in VE systems has received considerable attention. (See, for example, the study as well as the review of previous studies, in Watson, Walker, Ribarsky, and Spaulding, 1998.) Nevertheless, the amount of work on the effects of time delay on depth perception via motion parallax has been relatively limited. Previous work in this area (Satoh et al., 1991; McCandless et al., 2000) is considered in section 5 in conjunction with a discussion of our own results.

2 Experiment

The subject wore a Polhemus 3Space head tracker and was seated in front of a computer monitor 10.5 in. wide and 8.0 in. high at a distance (from eye to monitor) of roughly 25 in. In each trial of an experiment, the subject was presented with a visual stimulus on the monitor, moved his or her head laterally while looking at the stimulus, and then made a judgment about the nature of the stimulus.

All the stimuli used in the experiment were random-dot patterns that were structured in such a way that lateral movement of the head caused dots to move in a manner (via appropriate processing of the head movement information) that simulated the effects of viewing one cycle of a sinusoidal depth grating (figure 2). The grating was placed such that the variation in depth occurred along the vertical dimension, contours at constant depth lay in the horizontal dimension, and the horizontal contour corresponding to the midpoint of the vertical cycle lay in the plane of the monitor. Random-dot patterns (also used in the studies by Rogers and Graham, 1979, 1983, and by Ono et al., 1990) were chosen as the basic stimulus to minimize all depth cues except those associated with motion parallax.

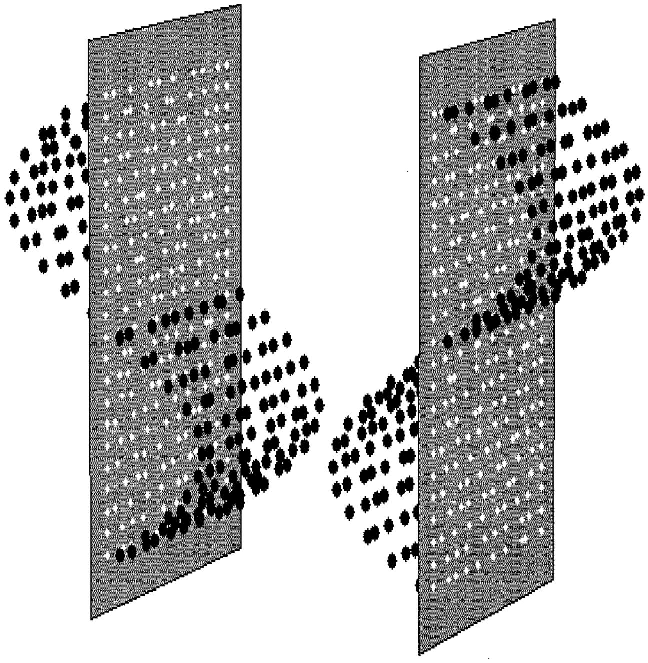

Figure 2. *Schematic illustration of the sinusoidal shapes that appeared when the subject's head moved laterally. The subject's task was to discriminate between sinusoids of phase 0 and phase π.*

Half of the stimuli were gratings with phase 0, and half were gratings with phase π. In other words, half of the stimuli were constructed so that, when the head was moved, the top half of the sinusoid appeared in front of the monitor plane (and the bottom half behind this plane), and half the other way. In all experiments, the task of the subject was to discriminate between these two classes of stimuli, that is, to judge whether the stimulus had phase 0 or phase π.

Each random-dot pattern was composed of approximately 150 dots. With the given monitor and seating arrangement, the dimensions of each dot (in terms of angle subtended at the eye) were roughly 14 min. wide

Table 1. *Conversion of Sinusoidal Amplitude A to angular disparity θ at the eye of the observer*

Amp. A (in)	0.01	0.08	0.16	0.32	0.64	1.00
Ang. Disp. θ (min)	0.23	1.86	3.72	7.39	14.59	22.48

and 5 min. high (the asymmetry resulting from processing performed to smooth the lateral motion of the dots associated with the lateral head movements). These dots were distributed over seventeen rows of dots on the display with adjacent rows separated by roughly 23 min. A lateral change of 0.1 in. in the position of a dot on the monitor corresponded to a change of roughly 14 min. in angle at the eye.

These random-dot patterns were programmed so that, when the subject's head moved laterally, the dots moved laterally, after a controlled time delay, in a manner that simulated the changing view that would be seen if the subject moved laterally in front of the above-described sinusoidally shaped surface.

The set of stimuli involved variation over two parameters in addition to the phase (0 or π) of the sinusoid: the amplitude A of the sinusoid and the time delay T between the head movement and the image update. A was chosen randomly without replacement from the set {0.01, 0.08, 0.16, 0.32, 0.64, and 1.00 in.}, and T was chosen randomly without replacement from the set {100, 155, 265, 485, 925, and 1805 ms}. (Inasmuch as A is defined as the amplitude of the sinusoid, the maximum excursion in depth over a whole cycle is given by 2A.) For each combination of A and T, the computer generated sixteen random-dot patterns. Thus, altogether, there were 6 × 6 × 16 = 576 stimuli available for presentation (half of phase 0 and half of phase π).

For the given geometric conditions, the maximum angular disparity θ at the eye of the observer corresponding to each of the sinusoidal amplitudes A tested is shown in table 1. The values of θ in this table give the disparities (in azimuth) that arise when points at the maximum or minimum of the sinusoid (which evidence the maximum horizontal movement when the head is moved laterally) are compared to the corresponding points halfway between the maximum and minimum of the sinusoid (which remain fixed when the head is moved) for a lateral off-center head position of 4.25 in. (the maximum off-center head displacement employed in the experiment, independent of A).

It should also be noted that the values of T specified above represent the mean values of T. The actual values (measured, using optical sensors, from the time at which the head started moving to the time at which the dots on the monitor started moving) varied randomly about the means with a standard deviation σ of roughly 20 ms independent of the mean value of T. This jitter in T was due primarily to the finite frame rates in the monitor and head tracker. (In the report by Yuan, 1999, not only was this jitter ignored, but the mean delays were incorrectly specified as 55, 110, 220, 440, 880, and 1760 ms.)

The experimental paradigm employed was a one-interval, two-alternative-forced-choice paradigm with equal a priori probabilities and without trial-by-trial feedback. In other words, on each trial, one stimulus was presented; this stimulus was drawn randomly with equal probabilities from the phase-0 class and the phase-π class (and with equal probabilities from the various A, T combinations except for the constraint of no replacement); the subject was forced to choose whether the stimulus belonged to the phase-0 class or the phase-π class; and the subject was not given any information after making the response about whether the given response was correct.

The different values of A and T were mixed together in the tests (as opposed to testing each combination of A and T separately) for two reasons. First, the variation over A and T from trial-to-trial minimized the possibility of subjects discriminating on the basis of cues other than the intended depth cues. (The fact that subjects actually based their responses on the intended depth cues was reinforced by subjective reports made by the subjects.) Second, the mixing together of the values of A and T in the tests reduced the possibility of introducing artificial dependence on A and/or T associated with learning or fatigue factors.

Two versions of the experiment were run, each on three MIT students. Except for the identity of the subjects (the two groups were disjoint except for one subject, the first author of this note), the only difference between the experiments was the following. In experiment 1, the subject was permitted to move his or her head from side to side for as long as he or she wished before making a response. In experiment 2, the subject was constrained to one-quarter of a head-movement cycle. Specifically, the subject was instructed to view the stimulus only while moving once from the middle to the left.

Further details concerning the experimental equipment and the programming of the dot-pattern movements can be found in Yuan (1999).

3 Results

The results of experiments 1 and 2 are shown in figures 3 and 4. Whereas the top two rows in each figure show how percentage correct (out of sixteen trials for each subject) depends on A with T as the parameter, the bottom two rows show how percentage correct depends on T with A as the parameter. The standard deviation of the percentage correct across subjects, averaged over all A and T, was roughly 11% for both experiments.

In viewing these figures, four points should be noted. First, the amplitude A of the sinusoidal depth pattern can be converted to angular disparity θ at the eye from the information provided in table 1. Second, the values of A and T are spaced logarithmically along the abscissa of the graphs in figures 3 and 4 (with the exception of the space between $A = 0.01$ and $A = 0.08$ along the A abscissa, which is arbitrary). Third, as indicated previously, the actual delays are random variables: the specified values of T correspond to the mean values of the random variables, and, in each case, the standard deviation σ is approximately $\sigma = 20$ ms. Fourth, in both figures 3 and 4, the data points obtained from the common subject in the two experiments (the first author of this paper) are shown by open circles.

Among the main results to be noted from the results shown in figures 3 and 4 are the following:

1. The value of T has essentially no effect on the discrimination performance for $T \leqslant 265$ ms. For both experiments, and all values of A, there is a negligible degradation in performance in going from $T = 100$ ms to $T = 265$ ms. Moreover, provided $A \geqslant 0.32$ in. ($\theta \geqslant 7.4$ min.), performance is relatively high throughout this range of T.
2. The amplitude threshold for performing the task at the smaller values of T (including, presumably, the untested value $T = 0$) appears to lie in the neighborhood of $A = 0.1$ in. ($\theta = 2.2$ min.). Results relevant to this conclusion are shown in table 2, where we have listed for each experiment (averaged over subjects) the values of percentage correct for $A = 0.01$, 0.08, and 0.16 in. and the two smallest values of T tested (100 and 155 ms). Although we are not able to make rigorous comparisons between this threshold and data in the literature, this threshold does not seem unreasonable in view of published results on the detection of movement or depth by motion parallax (For example, see Rogers and Graham, 1979, 1983.)
3. The results of experiment 1 show that, when the value of A is clearly above threshold, performance is clearly worse than chance at 925 ms. As can be seen in the bottom row of figure 3, for $A \geqslant 0.32$ in. ($\theta \geqslant 7.4$ min.), the percentage correct clearly dips well below chance level (50% correct) at $T = 925$ ms. In fact, averaged over the values $A = 0.32$, 0.64, and 1.00 in. for this experiment, the performance at 925 ms is only 16% correct. A similar (although somewhat less dramatic) dip below chance level is observed in the results of experiment 2 at a delay of 1805 ms. (See the bottom row of figure 4.)

It should also be noted when examining the data shown in figures 3 and 4 that two auxiliary tests were performed (on a single subject) to check procedures. In one test, experiment 1 was repeated with one eye closed. (For convenience, all the other tests were performed with both eyes open.) As expected, the results of this test showed no dif-

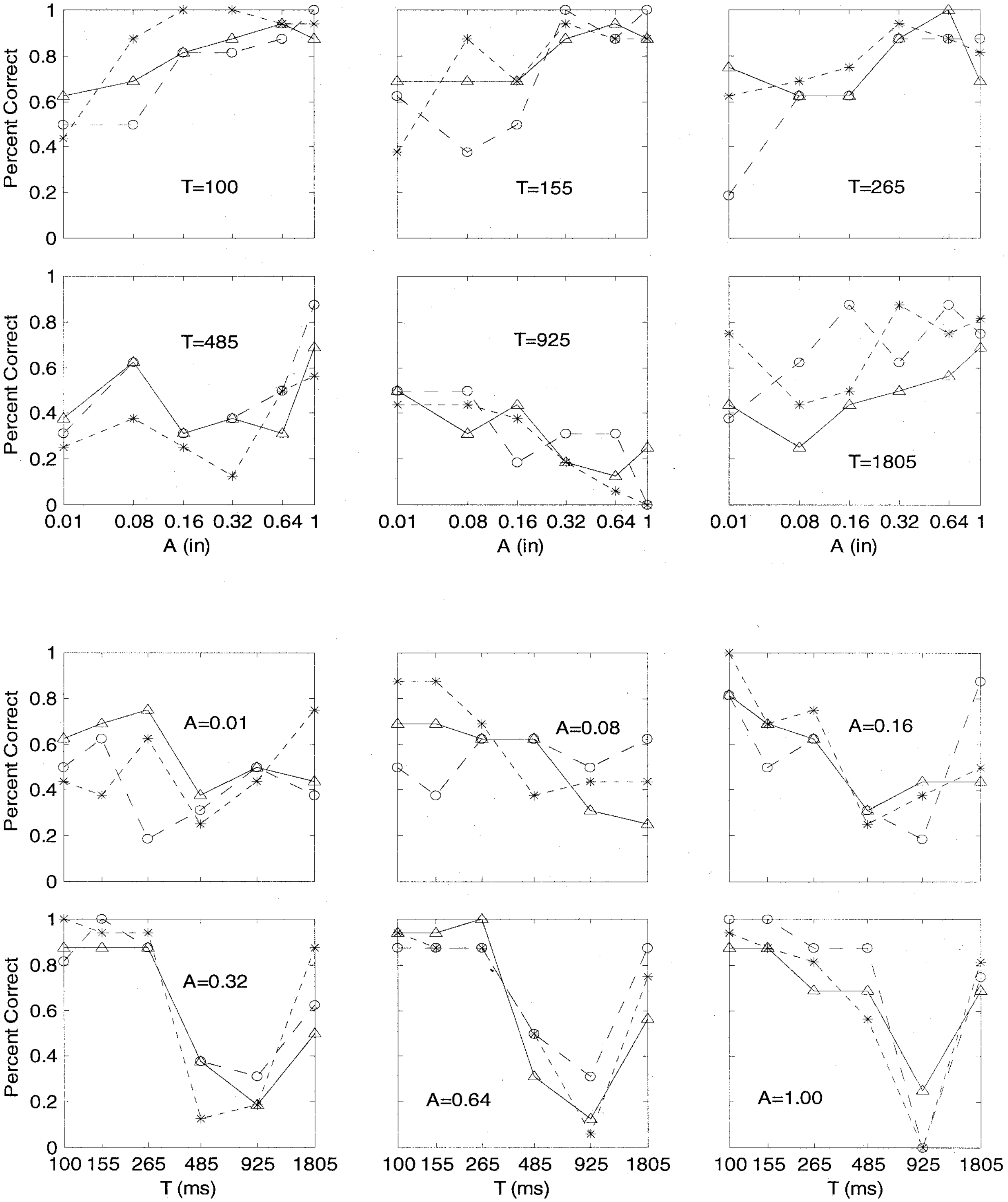

Figure 3. *Results of experiment 1. In all graphs, A is the amplitude of the sinusoid (in inches), and T is the mean time delay between head movement and image movement (in milliseconds). Note that the abscissa for both A and T are scaled logarithmically (except for the space along the A abscissa between A = 0.01 and A = 0.08, which is arbitrary). Averaged over all A and T, the standard deviation (around the mean) across the three subjects was σ = 11.6 % correct.*

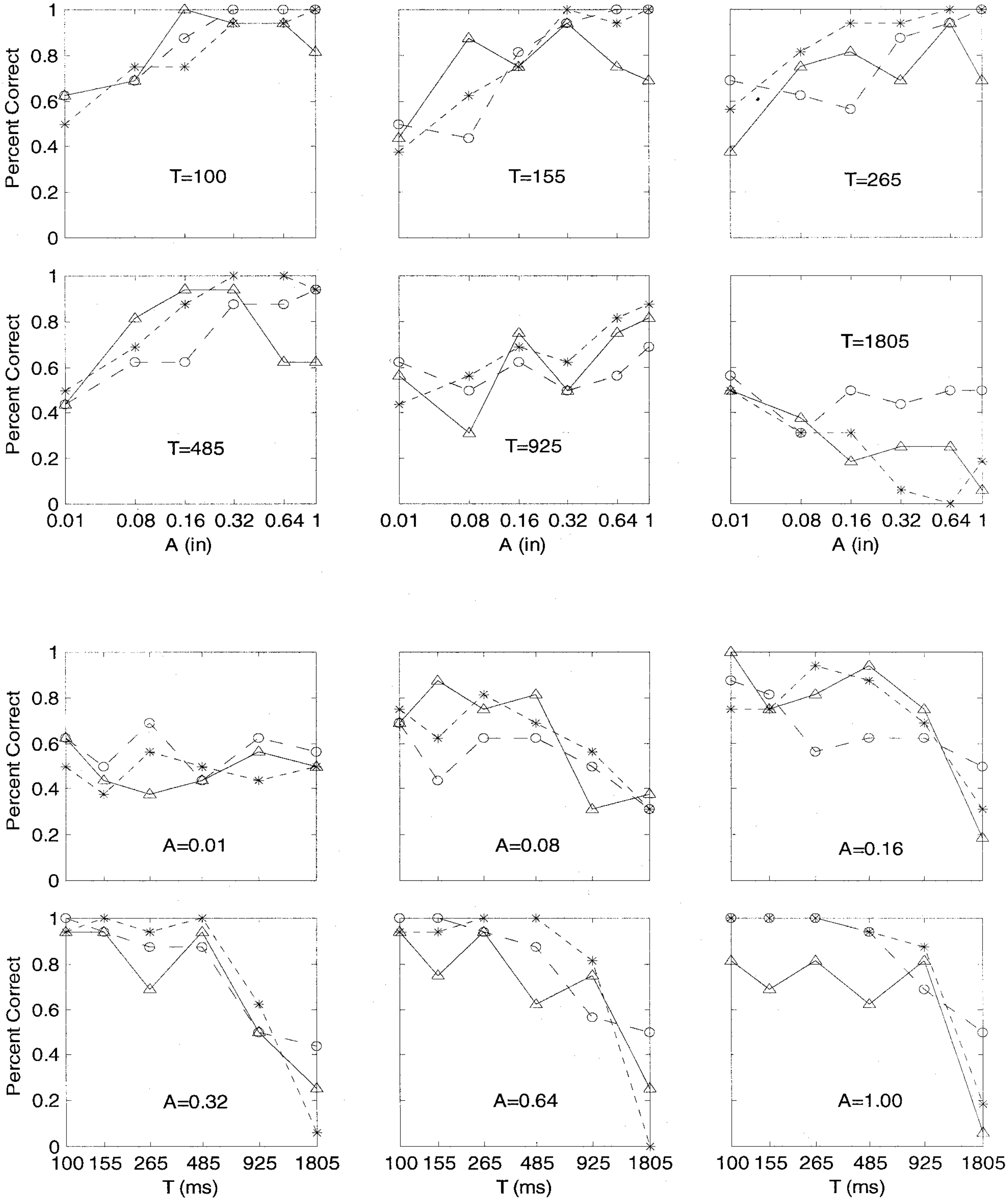

Figure 4. *Results of experiment 2. For this experiment, the standard deviation was* $\sigma = 11.1\%$ *correct.*

Table 2. *Percent Correct Scores (averaged over subjects) for T = 100 and 155 ms and A = 0.01, 0.08, and 0.16 in.*

Exp. 1			
	0.01	0.08	0.16
100	52%	69%	88%
155	56%	65%	63%
Exp. 2			
	0.01	0.08	0.16
100	58%	71%	88%
155	44%	65%	77%

ference in performance when the view was monocular. In the second test, the motion tracker was removed from the subject's head and the dots were caused to move by another person moving the tracker in a manner that could not be seen by the subject. Thus, the dots moved, but their motion was not correlated with any movements of the subject. As expected, performance under this condition was at chance level.

Further material relevant to observations 1 through 3 listed above are presented below.

4 Statistical Analysis

ANOVA analysis was performed on the percentage-correct responses to test the statistical significance of the trends observed in the data. For both experiments 1 and 2, a two-way ANOVA was performed to examine the significance of A, T, and their cross product A × T, on percentage correct (individual subjects were treated as repeated measures in this analysis). Because each of three subjects completed sixteen trials for each of 36 conditions in each of the two experiments, there were 108 separate estimates of percentage correct in each ANOVA.

In experiment 1, A $[F(5, 72) = 6.8, p \leqslant 0.0001]$, T $[F(5, 72) = 43.7, p \leqslant 0.0001]$, and their cross product A × T $[F(25, 72) = 3.6, p \leqslant 0.0001]$ yielded statistically significant effects. Similarly, in experiment 2, statistically significant differences were found for A $[F(5,72) = 12.3, p \leqslant 0.0001]$, T $[F(5, 72) = 40.4, p \leqslant 0.0001]$, and A × T $[F(25, 72) = 2.7, p \leqslant 0.0001]$.

For both experiments, post-hoc Bonferroni tests were conducted to determine exactly which conditions were significantly different at the $p = 0.02$ level.

In experiment 1, it appeared that performance generally improved with A. This observation was supported by the statistical analysis. Performance for the lowest-amplitude stimulus (A = 0.01 in.) was significantly worse than for the greatest three amplitudes (A = 0.32, 0.64, and 1 in.). No other amplitude differences yielded significantly different levels of performance, although the relatively good performance for the largest amplitude compared to intermediate amplitude conditions approached significance (that is, comparing A = 1 to 0.08 and 0.16 in. yielded p values of 0.05 and 0.06, respectively). Also in experiment 1, performance appeared to be invariant for delays between 100 and 265 ms, to degrade as delay was increased up to 925 ms, and then to improve as the delay was raised to 1805 ms. Again, these observations are supported by the statistical analysis. There were no significant differences in performance when comparing the three shortest delays (T = 100, 155, and 265 ms). However, delays of 485 and 925 ms yielded significantly lower performance than performance at all other delays (100, 155, 265, and 1805 ms). Performance with the longest delay (1805 ms) was significantly worse from that at the shortest delay (55 ms), and the difference in performance at 1805 versus 155 ms (the second-smallest delay) approached significance ($p = 0.024$). Finally, the difference in percentage correct for the two worst delays (485 and 925 ms) also approached significance ($p = 0.024$). Bonferroni tests of the cross-term differences were also conducted, and the only significant differences observed were consistent with the main effects already noted. For example, comparisons of conditions with a delay of 485 ms (and various amplitudes) to conditions with a delay of 100 ms (and various amplitudes) often yielded statistically significant effects.

In experiment 2, performance again appeared to improve with amplitude, and this pattern was again supported by the statistical analysis. Specifically, performance for the lowest amplitude stimuli (0.01 in.) was significantly worse than for the largest four amplitudes (0.16, 0.32, 0.64, and 1 in.), and performance for the second smallest amplitude stimuli (0.08 in.) was significantly worse than for the largest three amplitudes (0.32, 0.64, and 1 in.). No other amplitude differences yielded statistically significant effects. Unlike in experiment 1, in experiment 2, performance appeared to degrade when the delay was greater than 485 ms (and was relatively insensitive to delay at the smaller delay values). These observations are also supported by the statistical analysis. Specifically, performance with the longest delay (1805 ms) was significantly worse than for all other delays, and performance with the second-longest delay (925 ms) was significantly worse than for all the delays less than 925 ms (that is, 485, 265, 155, and 100 ms). No other differences across delays yielded statistically significant effects. As in experiment 1, when explicitly comparing all possible pairs of the 36 conditions, the only comparisons that yielded statistically significant effects were those that were consistent with the main effects (that is, performance for long-delay conditions was often significantly worse than performance for short-delay conditions).

5 Discussion

The occurrence of performance that is substantially worse than chance at large values of the delay T (item 3 in section 3) can be explained, at least in part, in terms of the particular head movements employed. In particular, the head movement recordings in experiment 1 show that the delay of 925 ms at which performance is substantially worse than chance is approximately equal to the duration of half a cycle of the quasiperiodic side-to-side head movements of the subject. Under such conditions, if the subject does not take explicit cognitive account of the delay T and operates as if there were no delay, any stimulus in the phase-0 class will appear to be in the phase-π class, and vice versa. Furthermore, the fact that the delay T was varied randomly from trial to trial would make taking such cognitive account of the T value exceedingly difficult.

This explanation for the worse-than-chance performances in experiment 1 at 925 ms also appears applicable to the worse-than-chance performance of two of the subjects (but not the third) in experiment 2 at 1805 ms. In the latter experiment, subjects were instructed for each trial to move their head from center to left, stop the head movement and shut their eyes, return their head to center, and then select a response based on their perception during their movement from center to left. If these instructions were not followed carefully and the subject's eyes were open during at least part of the return trip from left to center, and the response was influenced by what was seen during this return trip, then one might expect the same worse-than-chance performance to occur in experiment 2 (at least to some extent) as in experiment 1. This notion is supported by two facts. First, the leading author, who was extremely careful to follow the eyes-shut instruction, does not show worse-than-chance performance at 1805 ms. Second, when the time course of the head movements for experiment 2 was compared to that in experiment 1, it was discovered that the time required for the half-cycle movement in experiment 2 was roughly double that in experiment 1 (presumably due to the need to stop the head at the left-most position before embarking on the return to center).

The fact that the time required for the half-cycle movement in experiment 2 (figure 4) was roughly double that in experiment 1 (figure 3) also appears to be reflected in a comparison between the results shown in the top six graphs of figures 3 and 4 (that is, the graphs in which A is given along the abscissa and T as parameter). More specifically, note the similarity between T=925 ms (figure 3) and T=1805 ms (figure 4), between T=485 ms (figure 3) and T=925 ms (figure 4), and between T=265 ms (figure 3) and T=485 ms (figure 4). In other words, to a rough first-order approximation, one can obtain the graphs in figure 4 from those in figure 3 by essentially doubling T.

The principal result of these experiments—namely, that the delay T has no major effect on performance for

T ⩽ 265 ms (item 1 in section 3)—is basically good news. Although larger values of T might be expected in some systems and some situations, T should lie well within this range in most cases. Of course, the extent to which the results we have obtained with the random-dot patterns and the simulated sinusoid generalize to other situations remains to be seen.

Finally, although it is exceedingly difficult to make meaningful comparisons between our data and those reported previously in the literature, two points should be noted.

First, in the study by Satoh et al. (1991), in which the effect of horizontal head movement was compared to that of vertical head movement (for an "empty passage" image that was unbiased for the two dimensions), it was found that the maximum delay time for which subjects always rated (100%) the stimulus as showing little or no influence of delay time on motion parallax for the horizontal case was 180 ms. (The result when the 100% criterion was reduced to a 50% criterion was 300 ms.) Second, in the recently reported motion parallax results of McCandless, Ellis, and Adelstein (2000) on distance estimation of a virtual object superimposed on a real environment (using a see-through head-mounted display), the standard deviation of their distance judgments was essentially independent of time delay for delays in the region 31 to 197 ms. Although their primary interest was in the mean judgments rather than the standard deviation of these judgments, the standard deviation in their results is more closely related to the discrimination task considered in this paper.

If we have the opportunity to extend the work discussed in this research note, we intend (among other things) to record and analyze on a trial-by-trial basis the relationship between the subject's head-movement behavior (position, velocity, and acceleration) and the subject's response behavior.

Acknowledgment

This work was supported by ONR Grant N00014-97-1-0635. We are also grateful to the anonymous reviewers for their valuable comments on an earlier draft of this note.

References

Ferris, S. (1972). Motion parallax as a determinant of perceived depth. *Journal of Experimental Psychology*, *95*(2), 258–263.

Hoshino, H., Hiruma, N., Yamada, M., & Fukuda, T. (1991). Relation between visual effects and gain of motion parallax. *Systems and Computers in Japan*, *22*(3), 56–65.

McCandless, J. W., Ellis, S. R., & Adelstein, B. D. (2000). Localization of a time-delayed monocular virtual object superimposed on a real environment. *Presence: Teleoperators and Virtual Environments*, *9*(1), 15–24.

Ono, H., & Steinbach, M. (1990). Monocular stereopsis with and without head movement. *Perception & Psychophysics*, *48*(2), 179–187.

Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, *8*(2), 125–134.

———. (1983). Anisotropies in the perception of three - dimensional surfaces. *Science*, 221(4618), 1409–1411.

Satoh, T., Tomono, A., & Kishino, F. (1991). Allowable delay time of images with motion parallax, and high-speed image generation. *SPIE* 1606, 1014–1021.

Watson, B., Walker, N., Ribarsky, W., & Spaulding, V. (1998). Effects of variation in system responsiveness on user performance in virtual environments. *Human Factors*, *40*(3), 403–414.

Yuan, H. (1999). *Effects of delays on depth perception by motion parallax in virtual environments.* Unpublished master's thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, MA.