Low-Dimensional Representations of Visuomotor Coordination for Natural Behavior

A Preprint

Ashima Keshava
University of Osnabrück
Germany
akeshava@uos.de

Maximilian A. Wächter
University of Osnabrück
Germany

Franca Boße
University of Osnabrück
Germany

Thomas Schüler
Halocline GmbH & Co. KG
Germany

Peter König
University of Osnabrück
University Medical Center,
Hamburg-Eppendorf
Germany

Abstract

Understanding how the eyes, head, and hands coordinate in natural contexts is a critical challenge in visuomotor coordination research, often limited by sedentary tasks in constrained settings. To address this gap, we conducted an experiment where participants proactively performed pick-and-place actions on a life-size shelf in a virtual environment and recorded concurrent gaze and body movements. Subjects exhibited intricate translation and rotation movements of the eyes, head, and hands during the task. We employed time-wise principal component analysis to study the relationship between the eye, head, and hand movements relative to the action onset. We reduced the overall dimensionality into 2D representations, capturing over 50% of the explained variance and up to 65% just in time of the actions. Our analysis revealed a synergistic coupling of the eye-head and eye-hand systems. While generally loosely coupled, they synchronized at the moment of action, with variations in coupling observed in horizontal and vertical planes, indicating distinct mechanisms for coordination in the brain. Crucially, the head and hand were tightly coupled throughout the observation period, suggesting a common neural code driving these effectors. Notably, the low-dimensional representations demonstrated maximum predictive accuracy ~200ms before the action onset, highlighting a just-in-time coordination of the three effectors. This study emphasizes the synergistic nature of visuomotor coordination in natural behaviors, providing insights into the dynamic interplay of eye, head, and hand movements during reach-to-grasp tasks.

***K*eywords** low-dimensional representation, visuomotor coordination, eye-head-hand coordination, ecological validity, virtual reality

1 Introduction

Eye-hand coordination is a defining characteristic of everyday activities. Routine tasks involving object interactions require coordination between multiple sensorimotor systems to execute bodily transformations to effect environmental changes. Studies of tasks like sandwich-making (Hayhoe et al., 2003) or tea-making (Land et al., 1999) show how the visual information is sampled by the eye to direct and guide hand movements to accomplish goals iteratively until the task is adequately terminated. In this sense, eye movements serve a predominant function of planning and assisting manual interactions in the environments.

Task constraints, available sensory information and the cognitive context, can affect eye-hand coordination (Hu and Goodale, 2000; Hayhoe et al., 2003; Droll and Hayhoe, 2007). Studies have shown that gaze targets earmark upcoming sensorimotor events (Flanagan et al., 2006; Belardinelli et al., 2016). There is also consistent evidence that gaze control supports predictive motor control in object manipulation (Johansson et al., 2001). These findings have shown eye movements gather information from the environment proactively and in anticipation of upcoming manual action (Land and Furneaux, 1997; Land and Hayhoe, 2001).Keshava et al. (2024) showed that vision-for-action in naturalistic tasks can be accomplished with just-in-time representations where gaze fixations relay information for an action right before the action. Common actions, such as picking up objects, require only intermittent visual fixations to guide actions rapidly and efficiently. While eye-hand coordination has been studied in various task contexts, there are still gaps in the literature concerning how the structure of the external world might affect these coordination strategies.

Studies in visuomotor control have investigated eye hand-coordination in constrained settings where stimuli are shown on a computer screen in a limited space or in a head-fixed setup (Johansson et al., 2001; Bowman et al., 2009; Danion et al., 2021). Biguer et al. (1985) reported that with head-fixed experiment designs, the eye-in-head signal accuracy deprecates at larger angles. This has implications for how most sedentary experimental setups generalize to real-world human behavior. Moreover, the experimenters often cue the movements, and the subjects do not make self-generated proactive movements. In naturalistic tasks, eye, hand, and head movements must be coordinated together in a common coordinate system. Even in natural everyday tasks such as sandwich and tea-making, the tasks involved visuomotor coordination in a restricted plane such as on a table or a kitchen countertop. More importantly, these studies do not adequately exploit the structure of the natural world where gaze and reach movements coordinate for actions at different locations in the environment with respect to the body. In everyday life, humans have to coordinate eye and hand movements to reach objects from the floor, a top shelf, or to the right or left of the body. This study makes the case for an ecological perspective of eye-head-hand coordination where the variance of natural body movements is also accounted for.

Ingram and Wolpert (2011) have endorsed naturalistic approaches to understanding sensorimotor control and how these approaches provide a necessary adjunct to traditional lab-based studies. In recent years, virtual reality (VR) has facilitated the study of eye and body movement behavior in controlled naturalistic settings (Keshava et al., 2020, 2023; König et al., 2021). VR headsets and body trackers offer us the means to simultaneously record various oculomotor and body kinematic signals in a reliable way while subjects interact with the virtual world. Experiments in VR have grown popular in the last decade and have shown great promise in studying cognition in naturalistic and controlled environments.

As natural behavior is highly complex and varied, assessing the mechanism of coordination is not trivial. In natural contexts, we typically record high-dimensional data from multiple sensors. These three-dimensional ocular-kinematic signals comprehensively describe the translation and rotation behavior of the organism's many effectors. Hence, a low-dimensional representation of the multifaceted behavioral signals is of specific interest to characterize the diverse states of an acting animal (Bialek, 2022). Indeed, there is direct evidence of low-dimensional structures that depict distinct motor behaviors in humans (Santello et al., 1998; Sanger,

2000). In this regard, Principal Components Analysis (PCA) can emphasize the variation and reveal latent patterns in high-dimensional data while minimizing information loss. The computed low-dimensional subspace can be further explored to understand the underlying relationships of the original variables and their joint contributions to the variance in the data.

In this study, we explored human volunteers' movement trajectories while sorting objects on a 2m wide and 2m high life-size shelf in VR. They performed pick and place actions iteratively until a given sorting was achieved. Importantly, the participants were free to move, generate their own movements, and were not constrained by time. Consequently, they could move as naturally as possible in the virtual environment. The tasks were generic (in the sense of pick and place task) as well as novel (with regard to planning how to sort the objects) so that the findings could be generalized to visuomotor coordination in a proactive and natural context. As the movements made by subjects were complex in 3D space, which allowed for free translation and rotation movements, we explored the low-dimensional embedding of visuomotor coordination in pick-up and place actions. We were also interested in how the low-dimensional space accounted for the variance in the coordinated actions. Consequently, we examined the predictive power and timing of low-dimensional visuomotor control relative to action-critical events.

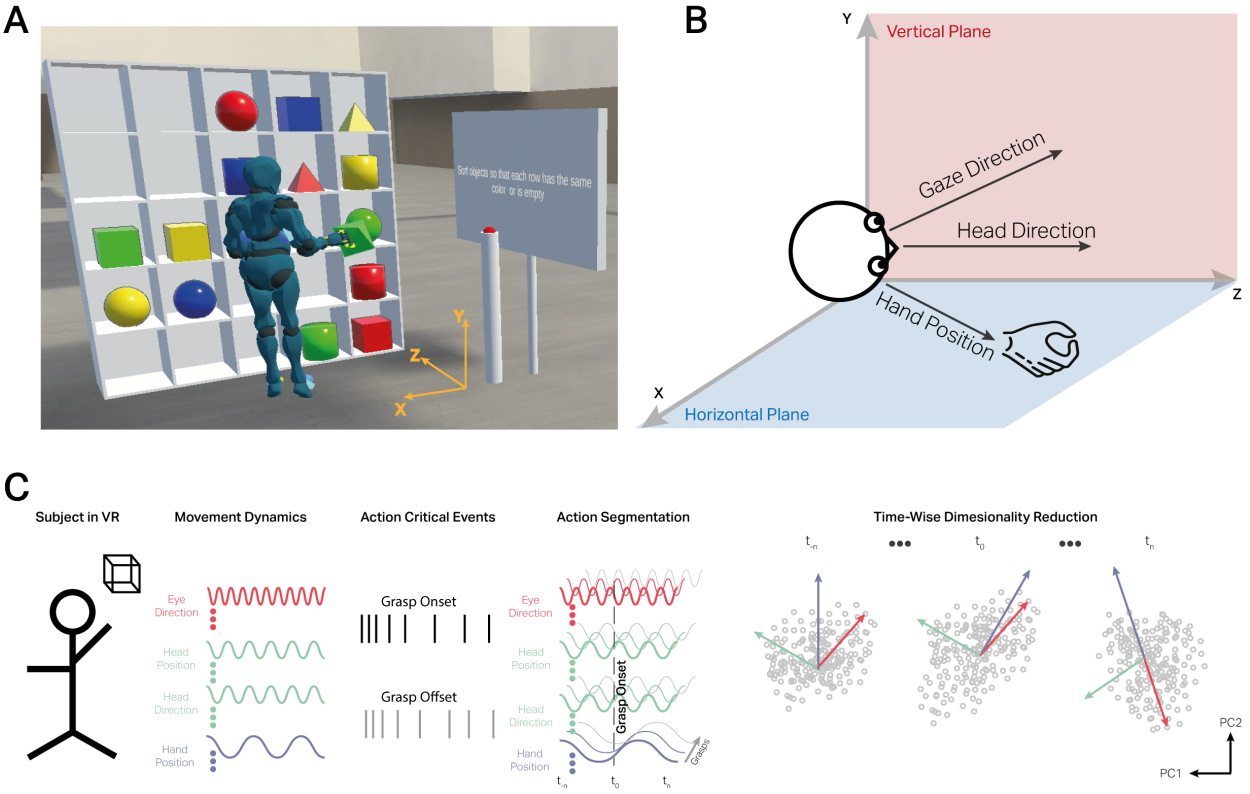

Figure 1: Experimental Setup. **A.** In a virtual environment, 27 participants sorted 16 different objects based on two features (color or shape) while we measured their eye and body movements. The objects were randomly presented on a 5x5 shelf at the beginning of each trial and were cued to sort objects by shape and/or color. All participants performed 24 trials in total with no time limit. **B.** We measured the position and the direction vectors of the head in a global reference frame and the eye direction and hand position in the head reference frame. **C.** To understand the coupling of the different sources of data from the head, eye, and hand in different reference frames, we performed dimensionality reduction analysis on the measured signals relative to action critical events of grasp onset (when the hand touches the object to pickup and the VR controller trigger is pressed) and grasp offset (when controller trigger is released and the hand drops the object on the desired shelf after object displacement). For each object displacement, we segmented the data to 1s before and after the action events. For each subject, the data matrix was composed of $G \times M \times T$, where G denotes object displacements, M denotes the 12 features (four data streams in x, y, z coordinates), and T denotes each time point relative to the action events. We applied Principal Component Analysis at each time point to the $G \times M$ matrix and reduced its dimensionality to $G \times 2$. The resulting 2D representation across time would then elucidate the contribution of the original data sources to the principal components and their explained variance.

2 Results

Twenty-seven healthy human volunteers performed an object-sorting task in VR on a 2m high and 2m wide life-size shelf. Participants performed 24 trials each. In each trial, participants sorted 16 randomly presented objects on the shelf based on a cued task and iteratively performed pick and place actions until the cued sorting was achieved. Figure 1A illustrates the experimental setup in VR. The experimental task consisted of sorting objects by their features. Each object was differentiated by color and shape. As the objects were randomly presented to the participants, they made reaching movements to different locations in space, such as reaching for objects closer to their feet or to the top of their heads. As participants performed these tasks, we simultaneously recorded their eye, head, and hand movement trajectories in 3D. In doing so, we recorded ecologically valid movements within a large spatial context.

For this study, we utilized four streams of continuous data. Namely, the head position of the participants, the unit direction in which the head was oriented, the cyclopean eye (average of left and right eye) unit direction within the head, and the hand position in world coordinates. These 3D movement vectors were represented in (x, y, z) coordinates (Figure 1B). In all, we analysed data from 27 participants, 5664 grasp actions with 209.77 ± 3.41 grasps per subject.

At the outset, we down-sampled the data to 40 Hz so that the continuous samples were equally spaced at 25ms. The down-sampling also smoothed the movement trajectories. We segmented the data streams relative to the onset of a grasp. We chose the time window from 1s before the grasp onset to 1s after. We similarly epoched the data relative to the grasp offset with the same window lengths. This allowed us to capture the eye, head, and hand movements while reaching the object, grasping it, and guiding it to a desired location. For both grasp onset and grasp offset events, we had 80 time points starting from -1s to 1s relative to an action, where time point 0 marked the onset or offset of said action. We then used PCA to reduce the overall dimensionality of the segmented data at each time point. The original data comprised of 12 dimensions (four data streams, each with 3d coordinates). We performed a time-wise PCA on each of the 80 time points to illustrate how the low-dimensional representation of the ongoing visuomotor coordination morphed relative to an action event and how the individual eye, head, and hand orientations contributed to these low-dimensional representations across time. Figure 1C depicts the data segmentation steps and PCA analysis approach in this study.

2.1 Complexity of Natural Behavior

Participants performed complex movements, which led to simultaneous translation and rotation of the head. To understand the range of movements made by the participants, we transformed the position and rotation vectors into 3D Euler angles (see section 5.4). Figure 2A shows the joint distribution of the initial position of the participants at the beginning of each trial in the horizontal and vertical planes. All participants started the trial from a fixed position in VR. The mean initial position of the participants in the horizontal plane was $0.0m, SD = \pm 0.03$ and in the vertical plane was $1.63m \pm 0.07$. The variance in the vertical direction corresponds to the variance in the height of the participants. From the initial positions, we calculated the translation-based deviations of the participants during the trials. Figure 2B shows the bi-variate distribution of the horizontal and vertical translation head movements from the initial position. In the horizontal plane, the mean deviation of the head was $-0.04m \pm 0.34$ and in the vertical plane it was $-0.07m \pm 0.14$. As evidenced by the spread of the distributions, participants made more translation movements in the horizontal plane. The downward movements show the translation of the head needed to interact with objects in the lower shelf locations. Figure 2C shows the bi-variate distribution of head rotation in the horizontal and vertical planes. The mean rotation in the horizontal plane was $-2.80° \pm 18.90$, and $-15.84° \pm 18.70$ in the vertical plane. The data shows participants made symmetrical horizontal head rotations. However, they had

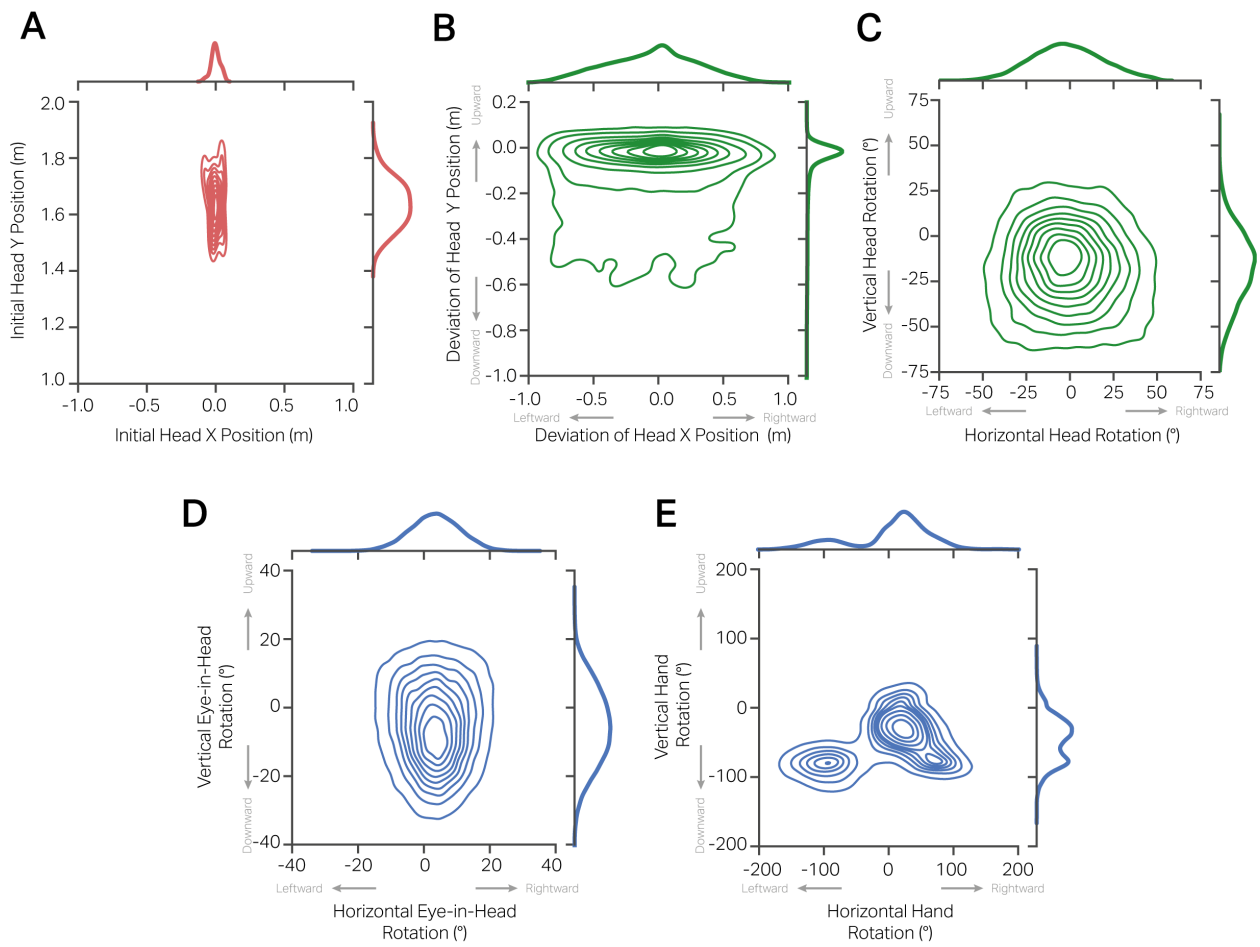

Figure 2: Complexity of natural movements. Plots show joint and marginal distributions of the participants' head, eye, and hand movements in the horizontal and vertical planes. **A.** Bi-variate distribution of head position at the start of each trial. **B.** Bi-variate distribution of the deviations of head movements from the initial position in the trials in the horizontal (abscissa) and vertical (ordinate) directions. **C.** Joint distribution of head rotations during the trials. When participants aligned their heads straight ahead facing the shelf, the rotation in the horizontal and vertical direction was 0°. **D.** Joint distribution of eye-in-head rotation during the trials. **E.** Bi-variate distribution of hand position in the head reference frame.

a tendency to direct their heads downward in the vertical plane. Taken together, the head movement vectors spanned the dimensions of the shelf. The distributions of the translation and rotation of the head showed no outlying behavior of the participants.

As above, we were interested in the range of motion exhibited by the participants' eye and hand movements during the trials. Figure 2D shows the bi-variate distribution of the eye rotation in the head reference frame. The mean eye rotation in the horizontal plane was $3.42° SD = \pm 7.50$. Similarly, in the vertical plane, the mean rotation was $-5.49° \pm 10.98$. Figure 2E shows the distribution of the hand rotation in the head reference frame. The mean rotation in the horizontal plane was $4.15° \pm 61.43$, and in the vertical plane, the mean rotation was $-46.12° \pm 31.79$. As seen in the data, the eye had symmetric rotations in both axes, with a larger variation in the vertical plane. Moreover, the average gaze orientation was slightly off-center with respect to the head orientation, i.e., within the head, the eye was slightly rightward and downward. This is likely an artifact of the task, as subjects were instructed to use their right hand to manipulate the objects.

The above data shows the complexity of natural behavior in a large spatial context. As the eye, head, and hand vectors show a complex structure, we could not isolate the individual horizontal and vertical components

of these vectors and directly correlate them. Hence, to decipher the associations between the eye, head, and hand orientations, we chose to study their relationship in a low-dimensional plane.

2.2 Low Dimensional Representations Visuomotor Coordination

Due to the complex range of motion exhibited by the eye, head, and hand, we studied the properties of the different movement vectors not as isolated systems but together. When the eye, head, and hand movement vectors are coordinated, we would find redundancies in the whole system, and fewer vectors could explain the joint coordination. As a first step, to understand the dynamic contribution of the different vectors relative to the start and end of a grasp, we performed a time-wise Principle Component Analysis (PCA). For each of these action-critical events, we performed the PCA using head position, head direction, eye direction, and hand position 3D vectors. We applied time-wise PCA to data from each subject. In doing so, we could ascertain first the average variance explained by the low-dimensional representations of the movement vectors across time and, subsequently, the contributions of the position and direction vectors to these low-dimensional representations.

Grasp onset

To understand the low-dimensional representation of eye, head, and hand movement vectors relative to grasp onset events, we performed time-wise PCA on head position, head direction, eye direction, and hand direction 3D vectors totaling 12 features per subject. Figure 3A shows the explained variance ratio (eigenvalues) of the principal components (PCs) across time, from 1s before action onset to 1s after. The explained variance of the top two PCs increases as the action onset approaches and subsequently reduces. At 1s before the action onset, the mean explained variance ratio for PC1 is $0.31(SD = \pm 0.03)$, and of PC2 is 0.18 ± 0.01. At time point 0, at the onset of grasp when the hand makes contact with the object to be picked up, PC1 and PC2 have a mean eigenvalue of 0.42 ± 0.03 and 0.22 ± 0.02, respectively. At the end of the grasp window, at 1s after the grasp onset, PC1 and PC2 have a mean eigenvalue of 0.33 ± 0.04 and 0.18 ± 0.01, respectively. Importantly, the peak explained variance ratio is achieved slightly before the action onset at $-0.19s, 95\%CI = [-0.24, -0.14]$. Throughout the grasp epoch, the other PCs exhibit an explained variance ratio of less than 0.15. This analysis shows that the 12 movement features can be reduced to a lower dimensional space consisting of two PCs that together explain more than 50% of the variance across time and are most informative near the action onset.

Next, to further understand the structure of the low-dimensional space, we plotted the contribution of each movement vector to this low-dimensional space. Figure 3B shows the projection of the data from an exemplar subject onto PC1 and PC2 for time points -1s before grasp onset, 0s at grasp onset and 1s after grasp onset. The direction of the component loadings illustrates its correlation with the first two PCs, and the length of the vector shows the magnitude of the correlation. The data are further colored according to the location where the upcoming pick-up action is performed. In this low dimensional representation, we see that 1s before the grasp onset the data begin to cluster according to the shelf height where the pickup action will subsequently happen. Furthermore, the loadings of the vertical (y) components of the head direction, eye direction, and hand position are similar and have a small angular deviation. The x components of these vectors have significant angular deviation and contribute differentially to the two PCs. At grasp onset (time=0s), the data are well clustered according to the pickup action location, and the horizontal and vertical components of the head direction, eye direction, and hand position point in the same direction. At time 1s after the grasp onset, the data are not clustered according to the pickup action location, and horizontal and vertical components of the movement vectors 'disengage.' The evolution of the PCA subspace across the entire time window is shown in Supplementary Material Movie 1. This analysis of the loadings of individual features shows their evolution within the low-dimensional space described by PC1 and PC2. Moreover, at grasp onset, the horizontal and

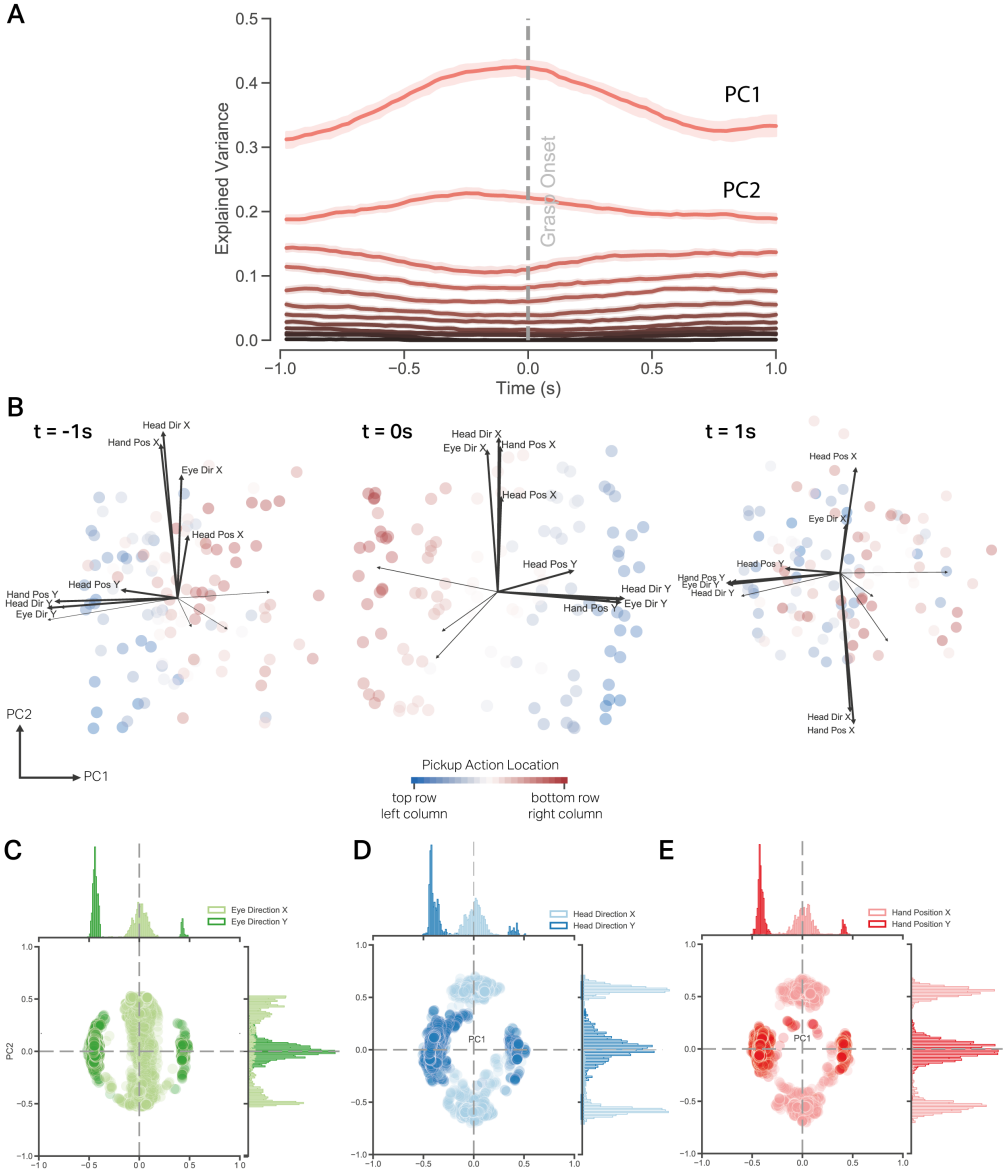

Figure 3: PCA analysis of 3D eye direction, head direction, head position, and hand position vectors. **A.** Temporal profile of eigenvalues of the 12 principle components (PCs) 1s before and after the grasp onset at time 0s. The solid lines denote the mean variance explained over the 27 participants by each PC, and the shaded region depicts the 95% confidence interval. **B.** Projection of exemplar subject data onto the first two PCs for time points -1s (left), 0s (middle), 1s (right). Each point within a time point denotes a pickup action projected onto PC1 and PC2, with the action colored according to the spatial location of the pickup action on the 5x5 shelf. The black arrow vectors represent the loadings of the original feature variables to the first two PCs. The magnitude of the vector denotes the strength of the correlation between the original variable and the two PCs. The direction provides the sign of the correlation (positive or negative). **C.** Bivariate distribution of the contributions of the horizontal and vertical eye direction vectors on the two PCs across the different time points and their respective marginal distributions. **D.** The joint contribution of the horizontal and vertical head direction vectors on the two PCs. **E.** The contribution of the horizontal and vertical hand position vectors on the two PCs across the different time points. As seen from these distributions, the horizontal components of the eye, head, and hand contributed primarily to PC2, and the vertical components contributed to PC1.

vertical components of the movement vectors are orthogonally represented, point in the same direction at grasp onset, and convey the same information to the two principal components.

Each principal component is composed of a linear combination of the original input variables. The weights of the original variables in the low-dimensional space indicate their respective correlations to the two PCs. The sign of the weights denotes the direction of the relationship with the PCs. To understand the overall contribution of the original movement vectors across all time points and participants, we charted their joint

Table 1: Mean absolute factor loadings of the original variables on the first two principal components across the grasping epoch

variable	PC1 Mean (SD)	PC2 Mean (SD)
Head Direction X	0.06 (0.05)	0.56 (0.07)
Head Direction Y	0.41 (0.03)	0.07 (0.05)
Head Direction Z	0.11 (0.08)	0.13 (0.10)
Head Position X	0.07 (0.05)	0.25 (0.12)
Head Position Y	0.29 (0.08)	0.08 (0.06)
Head Position Z	0.16 (0.10)	0.14 (0.10)
Eye Direction X	0.05 (0.04)	0.38 (0.13)
Eye Direction Y	0.43 (0.02)	0.06 (0.05)
Eye Direction Z	0.36 (0.07)	0.09 (0.08)
Hand Position X	0.06 (0.06)	0.54 (0.07)
Hand Position Y	0.40 (0.03)	0.05 (0.04)
Hand Position Z	0.38 (0.04)	0.10 (0.08)

contributions to the 2D eigenvectors (PC1, PC2) as shown in Figure 3C, D, E. The marginal distributions of the original variables in the principal component space show the vertical components of the eye, head direction, and hand position consistently contribute to the PC1, and the horizontal components correlate with the PC2. As seen from the plots, the horizontal components have a bimodal distribution on PC2 where equal proportions of participants' data contributed similarly to PC2 but in opposite directions at different points in time. Table 1 details the central tendency of the absolute value of the feature loadings in the low-dimensional space. Taken together, the PC1 and PC2 consisted of primarily vertical and horizontal components of the movement vectors, respectively, where the vertical components explained the largest proportion of the variance across time.

Grasp offset

Table 2: Mean factor loadings of the original variables on the first two principal components during grasp offset

variable	PC1 Mean (SD)	PC2 Mean (SD)
Head Direction X	0.07 (0.06)	0.57 (0.11)
Head Direction Y	0.42 (0.03)	0.08 (0.07)
Head Direction Z	0.12 (0.08)	0.14 (0.11)
Head Position X	0.06 (0.05)	0.19 (0.12)
Head Position Y	0.30 (0.08)	0.08 (0.07)
Head Position Z	0.17 (0.10)	0.14 (0.10)
Eye Direction X	0.06 (0.05)	0.36 (0.13)
Eye Direction Y	0.43 (0.02)	0.07 (0.06)
Eye Direction Z	0.34 (0.09)	0.13 (0.12)
Hand Position X	0.07 (0.06)	0.50 (0.11)
Hand Position Y	0.40 (0.03)	0.06 (0.04)
Hand Position Z	0.39 (0.04)	0.08 (0.07)

We repeated the above analysis for the object grasp offset events, i.e., when the object in hand is placed on the desired shelf. This was considered an action-critical event as the eye, head, and hand would have to coordinate to guide the object to a shelf and could be meaningfully different from the coordination required for reach movements. Figure 4A shows the explained variance ratio of twelve PCs when decomposing the head position, head direction, eye direction, and hand position vectors in 3D. Here again, the explained variance ratio of the first two PCs is around 0.50 throughout the time period and 0.60 close to the grasp offset. At timepoint -1s before the grasp offset, PC1 and PC2 had a mean explained variance ratio of 0.34 ± 0.03 and

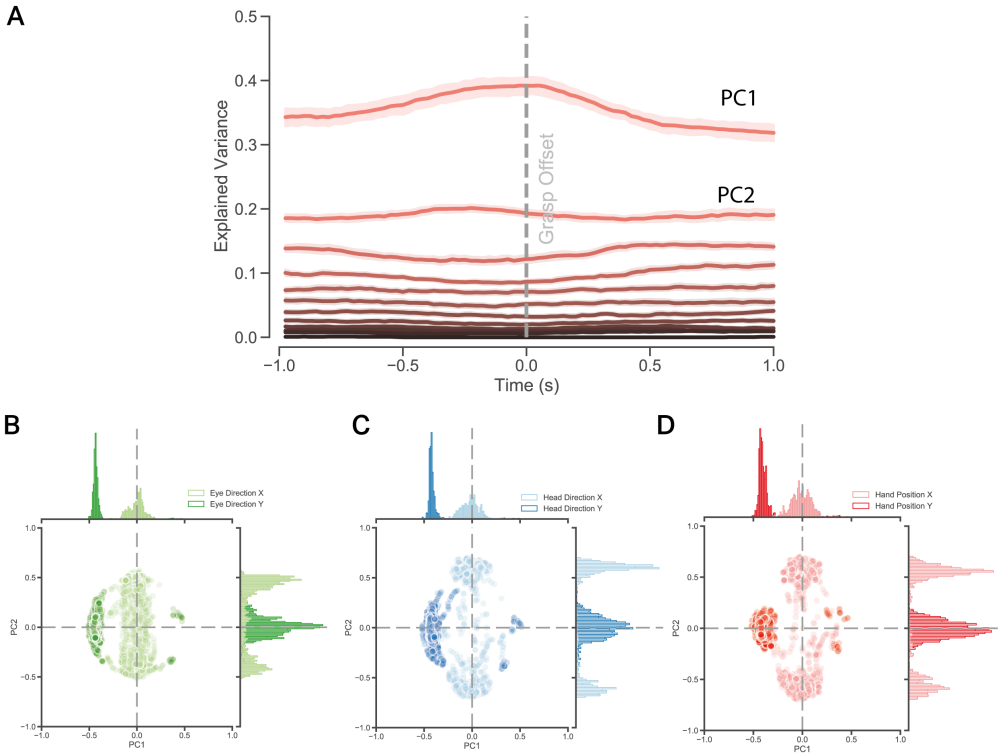

Figure 4: PCA analysis of 3D eye direction, head direction, head position, and hand position vectors. **A.** Temporal profile of eigenvalues of the 12 principle components (PCs) 1s before and after the grasp onset at time 0s. The solid lines denote the mean explained variance ratio over the 27 participants by each PC, and the shaded region depicts the 95% confidence interval. **B.** Bivariate distribution of the contributions of the horizontal and vertical eye direction vectors on first two PCs across the time points and participants, and their respective marginal distributions. **C.** The joint contribution of the horizontal and vertical head direction vectors on the two PCs. **E.** The contribution of the horizontal and vertical hand position vectors on the two PCs across the different time points. As seen from these distributions, the horizontal components of the eye, head, and hand contributed primarily to PC2, and the vertical components contributed to PC1.

0.18 ± 0.01. At time point 0, the mean explained variance ratio for PC1 was 0.39 ± 0.03 and 0.19 ± 0.01, respectively. Finally, at time point 1s after the grasp offset, PC1 and PC2 exhibited a mean explained variance ratio of 0.32 ± 0.03 and 0.19 ± 0.02, respectively. Similar to the grasp onset event, the increased explained variance ratio for the first two PCs indicates a higher correlation between the eye, head, and hand vectors. Nonetheless, the explained variance ratio of the first two PCs sums to 0.58 compared to the grasp onset events, which is at 0.64. This lower explained variance ratio at the grasp offset events suggests a lower correlation between the movement vectors.

As before, we checked for consistency of the contributions of the different movement vectors to the low-dimensional space. As seen from the joint distribution of the loadings of the original variable (Figure 4B, C, D), the horizontal components of the movement vectors consistently contributed to PC2, and the vertical components to PC1 across different time points and participants. Table 2 details the mean and standard deviation of the absolute contribution of each variable on the first two PCs. In sum, the low-dimensional space consisted primarily of the horizontal and vertical components of the movement vectors, where the vertical components explained more variance in the data than the horizontal components.

2.3 Cosine Similarity Analysis

In the above section, we performed time-wise PCA on action-critical events of grasp onset and grasp offset using head position, head direction, eye direction, and hand position vectors. The results showed a larger explained variance of the first two PCs, where this ratio was maximum just before the action events. PCA

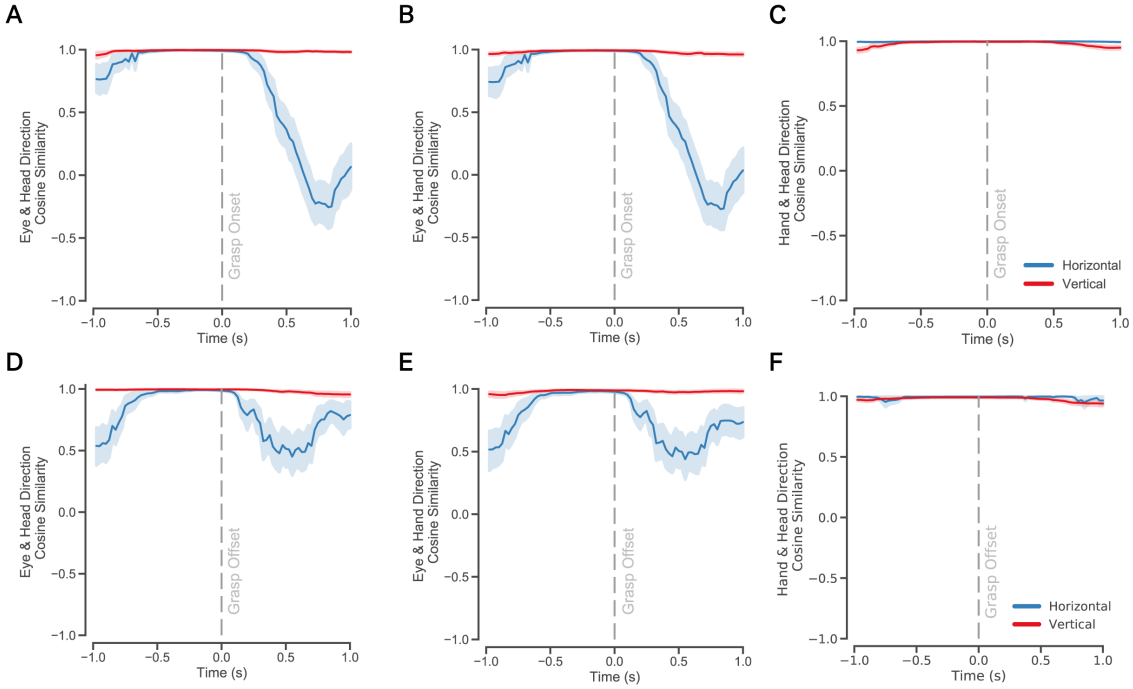

Figure 5: Cosine similarity between head, eye, and hand direction vectors for grasp onset and grasp offset events. **A.** Cosine similarity between the horizontal and vertical components of head and eye direction 3D vectors during grasp onset events. The solid lines denote the mean cosine similarity across 27 participants, and the shaded region denotes the standard error or mean. A cosine similarity value of 1 indicates a perfect correlation between the vectors in the PCA subspace, whereas a value of -1 denotes a perfect negative correlation. A value of 0 indicates the vectors are not related to each other. **B.** Cosine similarity between the horizontal and vertical components of eye direction and hand position 3D vectors during grasp onset events. **C.** Cosine similarity between the horizontal and vertical components of head and eye direction vectors during grasp offset events. **D.** Cosine similarity between the horizontal and vertical components of eye direction and hand position vectors during grasp offset events.

provides a subspace where the data's variance is maximized across the principal components. Using a vector similarity analysis, we can further explore relationships between the original variables within this subspace and identify which variables contribute similarly to the explained variance. Hence, exploration of the subspace can lead to a deeper understanding of the data structure, such as revealing groups of variables that might be part of the same underlying process or phenomenon.

To quantify the source of the increase in the explained variance ratio close to the action onset and offset events, we used cosine similarity analysis. Cosine similarity provides a measure of the correlation between the factor loadings in the 2D PCA subspace. Since PCA aims to identify patterns of similarity and differences across variables by transforming them into PCs based on their covariance, using cosine similarity to analyze further the orientation and correlation of variables in this transformed space complements the goals of PCA. It helps in understanding the structure and relationships between variables beyond mere dimensionality reduction. When the PC loadings point in the same direction, they indicate a high positive correlation with each other. Similarly, when they point in opposite directions, they indicate a high negative correlation. When they are orthogonal to each other, they indicate no correlation at all. Hence, using cosine similarity, we could ascertain the evolution of the correlation between the horizontal and vector components of the movement vectors before and after the action-critical events.

In grasp onset epochs, we calculated the cosine similarity between the x (horizontal) and y (vertical) components of the eye, head, and hand direction vector for each time point across subjects. Figure 5A shows the mean cosine similarity of the horizontal and vertical components of the head and eye direction vectors and the standard deviation. At time point 0, the horizontal and vertical components had a mean similarity of 0.99 ± 0.002 and 0.99 ± 0.001, respectively. Between eye and hand factor loadings (Figure 5B),

we observed a mean similarity of 0.99 ± 0.002 in the horizontal direction and 0.99 ± 0.001 in the vertical direction at grasp onset. Between head and hand factor loadings (Figure 5C), we observed a remarkable consistency throughout the action epoch where the similarity of the horizontal components was 0.99 ± 0.0003 and for vertical components 0.99 ± 0.001 at time 0s. Taken together, the vertical components of the eye, head, and hand factor loadings were well correlated throughout the grasp onset epoch. However, the horizontal components were correlated around 0.5s before the grasp onset, and this correlation reduced drastically shortly after the grasp event was triggered. Throughout the time course, the hand and head vectors varied in the same direction, both in the horizontal and the vertical planes, and exhibited a strong correlation.

We repeated the above analysis for the grasp offset events. First, we calculated the cosine similarity between the horizontal and vertical components of the eye, head, and hand factor loadings in the PCA subspace. Figure 5D illustrates the average similarity of eye-head loadings over subjects across the different time points. At time point 0, the average cosine similarity was 0.98 ± 0.004 in the horizontal direction and 0.95 ± 0.01 in the vertical direction. Between eye-hand factor loadings (Figure 5E), the horizontal and vertical components were perfectly aligned at time 0s and showed an average similarity of 0.97 ± 0.008 and 0.98 ± 0.003, respectively. Between head-hand factor loadings (Figure 5E), we observed a striking similarity as before, where the horizontal components had a mean similarity measure of 1.00 ± 0.003 and 0.99 ± 0.003 for the vertical components. Taken together, there was a strong coupling between the vertical components of the eye, head, and hand. In contrast, the horizontal components were aligned in the same direction briefly before the grasp offset. Further, the head direction and the hand position vectors covaried in the same direction and showed very high correlations.

The exploration of the PCA subspace with the loadings of the original variables showed interesting aspects of visuomotor coordination. Namely, the vertical components of the eye, head, and hand vectors were almost perfectly aligned in the low-dimensional space. The horizontal component of the eye direction vector, on the other hand, was only briefly oriented in the same direction as the head direction and hand position vectors at about 0.5s before the grasp onset and offset. This window of complete alignment of the vectors also coincides with the increase in the explained variance ratio of the first two PCs before the action onset. Crucially, the head direction and the hand position vectors tracked in complete unison throughout the action epochs. Thus, the similarity analysis of the effectors in the PCA subspace showed distinct coordination mechanisms for the horizontal and vertical components of the visuomotor system.

2.4 Generalization and Predictive Accuracy of the Low-dimensional Space

To further expound on the generalizability and predictive power of these low-dimensional structures, we predicted the location of the action at each time point based on the PCA-transformed data. For each time point t, we pooled the data from $N-1$ subjects and standardized it to zero mean and unit standard deviation. We then reduced the dimensionality of the data at each time point into a 2D space. For each time point, we trained a kernel-based support vector machine (SVM) to classify the location of the upcoming action. To achieve generalizability, we used leave-one-subject cross-validation. For each time point, we standardized the test data and transformed it using the PCA weights from the training data. We then computed the prediction accuracy of the trained SVM on the test data at that time point. We performed the above steps until each of the 27 subjects' data was used as test data for all time points ranging from 1s before and after the grasp onset. We repeated this analysis for the grasp offset events as well. Our analysis provided an aggregated prediction accuracy of the PCA-transformed training and test datasets. Thus, using cross-validation, we could generalize the information encapsulated in the PCA subspace and ensure our analysis was not affected by the peculiarities of single subjects.

The prediction of the object pickup action location across the grasp onset epoch provided a greater understanding of the evolution of the PCA subspace as seen in Figure 6A. At the 1s before grasp onset, the

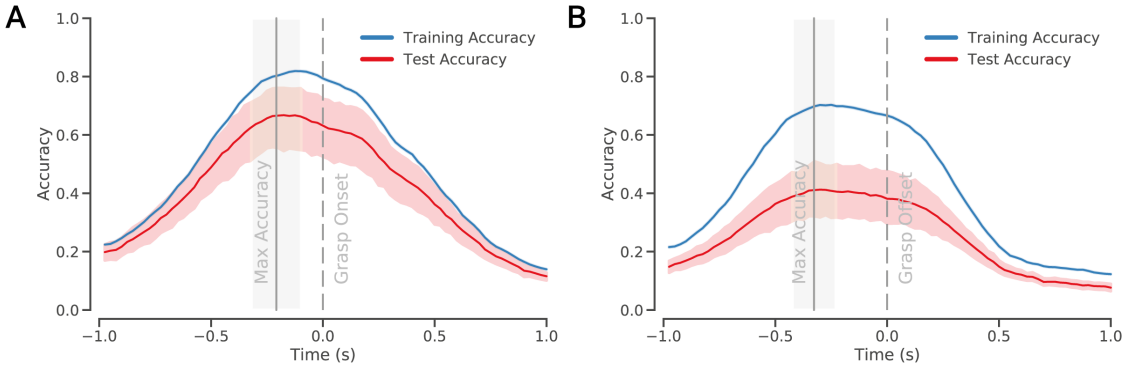

Figure 6: Accuracy of 2D PCA subspace in predicting the action location at each time point of the action epoch.**A.** The

prediction accuracy on the test data is low at $Mean = 0.20, 95\%CI = [0.17, 0.23]$. At time 0, the accuracy increases to 0.62 [0.52, 0.72]. At 1s after the grasp onset, the prediction accuracy is reduced to 0.12 [0.10, 0.13]. This shows that the PCA subspace encodes more relevant information about the upcoming action close to the action onset. Interestingly, the maximum predictive accuracy was not at the moment of grasp at time 0, but slightly earlier at -0.20s [-0.32, -0.09]. This indicates that maximum information about the coordination of the eye, head, and hand is available just in time for the action.

Similarly, the prediction of the object dropoff action locations across the grasp offset epoch is shown in Figure 6B. At 1s before the grasp offset, the predictive accuracy on the test data is 0.15 [0.13, 0.17]. At time point 0, the accuracy increases to 0.38 [0.30, 0.46] and decreases to 0.08 [0.06, 0.09] at 1s after the grasp offset event. Here again, the maximum predictive accuracy was at time point -0.32s [-0.42, -0.23] before the grasp offset. This is further indication that perfect coordination, which brings together all the components of the eye, head, and hand vectors into alignment for action, is achieved just in time.

With the above analysis, we generalized the performance of the 2D PCA subspace. We determined the predictive information encapsulated in the subspace and the timing of the best prediction. For both grasp onset and grasp offset, the PCA space's predictive accuracy increased with the approaching event and decreased after. Moreover, the maximum accuracy is achieved just in time for the action event. This can be accounted for by the large explained variance ratio of the 2D subspace and the high correlations between the eye, head, and hand orientation vectors in the subspace at that moment.

3 Discussion

The present study explored the low-dimensional representations of natural visuomotor coordination. Subjects exhibited complex translation and rotation movements with their eyes, head, and right hand by making reaching movements to pick up and place objects on a life-size shelf in VR. We applied a time-wise PCA on the position and orientation vectors of the different effectors to capture the explained variance at each time point relative to grasp onset and offset events. Our analysis showed the complex system composed of the eye, head, and hand could be well described in a 2D PCA subspace. The PCA subspace showed an increase in the explained variance ratio at grasping events (onset and offset), where more than 60% of the variance is accounted for by the first two eigenvectors. Our analysis demonstrates a dynamic coupling of the horizontal and vertical components of effectors just in time for the upcoming action. Furthermore, this coupling showed high predictive accuracy of the location of action. Hence, a dynamic coupling and decoupling of the movement vectors in the low-dimensional space exemplifies the synergistic role of the eyes with respect to the head and hand.

Methodological Considerations

Eye-hand or eye-head coordination is usually studied under constrained settings. In most cases, the horizontal and vertical positions or directions of the eye, head, and hand are extracted and directly correlated. In natural behavior, the complexity of the system does not afford such simplistic measures, as variables can interact with each other in non-obvious ways. Our study explored the latent relationships between the eye, head, and hand orientations in a low-dimensional space, helping to understand the underlying structure and relationships in the data relative to the action-critical events. Our aim was to capture both the explained variance and the evolution of the variable loadings across time.

Our analysis of the cosine similarity of the original variables in the PCA subspace revealed strong associations between the horizontal and vertical components of the eye, head, and hand vectors. PCA transforms variables into principal components based on their variances and covariances. When using cosine similarity to assess relationships between original variables based on their loadings, it's crucial to note that PCA mainly focuses on explaining variance, not necessarily revealing direct correlations between variables. It's important to clarify that cosine similarity measures the angle between two vectors and is a measure of orientation similarity rather than a direct measure of statistical correlation in the traditional sense (Pearson's correlation). Cosine similarity is less sensitive to the magnitude of vectors and focuses on their direction, which could be a limitation when the scale or variability of the original variables is relevant to their interpretation. In order to avoid drawing improper inferences, we plotted the distribution of the variable loadings on the first two PCs. We can confirm that the horizontal and vertical components of the eye, head and hand had similar magnitude of loadings on the PCs. Hence, by comparing the directionality of loadings, we could identify which variables share similar directional influences on the PCs, indicating underlying correlations that are not immediately obvious from the PCA results alone.

Finally, given the present study's naturalistic setting, various noise sources could affect the findings. The source of noise could be the eye or body trackers, which could exhibit errors due to slippage (Niehorster et al., 2020) or calibration errors (Ehinger et al., 2019). We calibrated the trackers after every three trials to mitigate such errors. Moreover, as participants performed the task while wearing the VR head-mounted display, the head movements could have been cumbersome when picking up objects from the lower shelf locations. We did not direct participants to move in any one particular manner and asked them to make movements that were comfortable for them. Nonetheless, the insights offered by our study open the door to further experimental replications to validate our findings.

Synergistic Coupling of the Effectors

The structure of the PCA subspace revealed a curious behavior of the eye, head, and hand direction vectors. Across grasping epochs, the vertical components of the effectors primarily contributed to the first PC. Also, the vertical components of the effectors were directionally aligned, as shown by the cosine similarity analysis. Thus, the vertical components of the eye, head, and hand varied in the same direction and contributed significantly to the overall variance explained. Conversely, the horizontal component of the eye direction vector aligned with the head and hand horizontal components shortly before the action onset. Land (1992) showed that head and eye movements are generated by receiving the same motor commands at almost the same time. Here, head movements are necessary to center gaze in the orbits. Our data implies that vertical movement is facilitated with head movements, and the eye completes the last leg of the operation by making horizontal adjustments. Hence, the vertical components of the eye and head direction vectors varied similarly, whereas the horizontal components showed a degree of independence.

Moreover, the horizontal and vertical components of the head and hand were aligned throughout the grasping epoch. Arora et al. (2019) corroborated this strong coupling between the head and arm with respect to the

eye in unrestrained macaque monkeys. Smeets et al. (1996) showed similar behavior during reaching tasks in humans where the head and hand reaction times and peak velocities were strongly correlated. Arora et al. (2019) and Smeets et al. (1996) argue that head movements facilitate foveation on the target to guide the final stages of object manipulation, leading to large correlations between the head and hand movements. In a similar vein, Pelz et al. (2001) showed a strong linkage between the head and hand movement trajectories, while the eye has a synergistic relationship instead of an obligatory one with them. Hadjidimitrakis (2020) hypothesizes that this strong head-arm coupling is a result of learned motor behaviors during feeding where the head and hand orientations are coordinated to bring food to the mouth. Head and hand coordination is not commonly studied, our results suggest the strong coupling between the two must be a consequence of a common neural code that drive this behavior.

In the present study, subjects reached for target objects while making large horizontal and vertical translation movements. This required them to take different body postures to accomplish reach movements to the top of their heads or near their feet. Our findings generalize the visuomotor coordination for different reach locations. Stamenkovic et al. (2018) showed that gaze, head, and hand synergies do not differ substantially under postural constraints, and the central nervous system adopts a whole body strategy subservient to achieving gaze stabilization. In this regard, head movements transform the eye-centered reference frame to a proprioceptive, hand-centered reference frame necessary for guiding hand movements. Thus, the PCA approach does not lose vital information about the covariance of the eye-head and eye-hand coordination in the presence of postural differences.

To validate the generalization of the PCA subspace, we predicted the grasp onset and offset locations by transforming the test data with the PCA weights of the training data. During reaching to grasp epochs, the prediction accuracy across time on the test data is similar to the training data. However, the prediction accuracy on the test data is considerably lower for grasp offset events. This is indicative of an idiosyncratic coordination required to guide objects and drop them to desired locations. This result is contrary to the findings of Pelz et al. (2001), where the gaze was always maintained on the object until the drop-off was complete. These differences probably arise due to differences in task design. In our study, subjects proactively picked up and dropped objects without a cued goal. Whereas Pelz et al. (2001) studied the pickup and drop-off of blocks to build a cued model. Our findings show pickup actions need precise visual guidance just in time of the grasp onset. However, object dropoff is achieved with mostly proprioceptive inputs with little guidance from the eye. Hence, the lower accuracy of the test set during grasp offset events is likely a consequence of individualized coordination between the effectors.

Neural Correlates of Multi-Effector Coordination

The high correlation between different effectors in a natural context is an indication of a common neural code that mediates visuomotor coordination. Buneo et al. (2002) showed that the posterior parietal cortex (PPC) and dorsal area 5 in macaque monkeys code the reach target location with respect to both eye and hand. They suggest that PPC achieves this transformation "by vectorially subtracting hand location from target location, with both locations represented in eye-centered coordinates." Similarly, in humans, there is growing consensus that the intraparietal sulcus (IPS) and anterior intraparietal area (AIP) are the seat of visually guided reach and grasp movements (Culham et al., 2006). Furthermore, the frontal eye fields (FEF), aside from controlling gaze shifts (Tu and Keating, 2000), may have a role in independent head control and is located adjacent to the dorsal premotor cortex (PMd), where neurons associated with both oculomotor and hand movement activities have been identified (Fujii et al., 2000). Arora et al. (2019) have further suggested that multi-effector coordination is facilitated in larger parietofrontal circuits. Thus, visuomotor coordination is achieved by a complex interplay between several brain regions and our study invites further research on this dynamic neural encoding of multi-effector coordination.

4 Conclusion

In this paper, we studied the low-dimensional representations of visuomotor coordination in natural behavior. The multidimensional data comprising of eye, head, and hand movement vectors could be decomposed to 2D representations that explained more than 60% of the variance in the data. A closer look at the subspace structure showed that the head and hand movements had a strong positive correlation and contributed substantially to the explained variance across time. However, the eye-head and eye-hand movements had distinct correlations in the horizontal and vertical axes. Moreover, the head and hand were tightly coupled throughout the observation period. These results show separate mechanisms of coordination where the head and hand are coordinated simultaneously, and the eye is coordinated for goal completion synergistically.

5 Methods

5.1 Participants

27 participants (18 females, mean age = 23.9 ± 4.6 years) were recruited from the University of Osnabrück and the University of Applied Sciences Osnabrück. Participants had normal or corrected-to-normal vision and no history of neurological or psychological impairments. They either received a monetary reward of € 7.50 or one participation credit per hour. Before each experimental session, subjects gave their informed consent in writing. They also filled out a questionnaire regarding their medical history to ascertain they did not suffer from any disorder/impairments that could affect them in the virtual environment. Once the consent was obtained, we briefed them on the experimental setup and task. The Ethics Committee of the University of Osnabrück approved the study (Ethik-37/2019).

5.2 Apparatus & Procedure

For the experiment, we used an HTC Vive Pro Eye head-mounted display (HMD)(110° field of view, 90Hz, resolution 1080 x 1200 px per eye) with a built-in Tobii eye-tracker[1] with 120 Hz sampling rate. With their right hand, participants used an HTC Vive controller to manipulate the objects during the experiment. The HTC Vive Lighthouse tracking system provided positional and rotational tracking and was calibrated for a 4m x 4m space. We used the 5-point calibration function provided by the manufacturer to calibrate the gaze parameters. To ensure the calibration error was less than 1°, we performed a 5-point validation after each calibration. Due to the study design, which allowed many natural body movements, the eye tracker was calibrated repeatedly during the experiment after every 3 trials. Furthermore, subjects were fitted with HTC Vive trackers on both ankles, both elbows and one on the midriff. The body trackers were also calibrated subsequently to give a reliable pose estimation using inverse kinematics of the subject in the virtual environment. We designed the experiment using the Unity3D[2] version 2019.4.4f1 and SteamVR game engine and controlled the eye-tracking data recording using HTC VIVE Eye-Tracking SDK SRanipal[3] (v1.1.0.1).

The experimental setup consisted of 16 different objects placed on a shelf of a 5x5 grid. The objects were differentiated based on two features: color and shape. We used four high-contrast colors (red, blue, green, and yellow) and four 3D shapes (cube, sphere, pyramid, and cylinder). The objects had an average height of 20cm and a width of 20cm. The shelf was designed with a height and width of 2m with 5 rows and columns of equal height, width, and depth. Participants were presented with a display board on the right side of the shelf where the trial instructions were displayed. Subjects were also presented with a red buzzer that they could use to end the trial once they finished the task. The physical dimensions of the setup are illustrated

[1] https://enterprise.vive.com/us/product/vive-pro-eye-office/
[2] Unity, www.unity.com
[3] SRanipal, developer.vive.com/resources/vive-sense/sdk/vive-eye-tracking-sdk-sranipal/

in Figure 1A. The horizontal eccentricity of the shelves from the center of the shelf extended to 89.2 cm in the left and the right directions. Similarly, the vertical eccentricity of the shelf extended to 89.2 cm in the up-down direction from the center-most point of the shelf. This ensured that the task setup was symmetrical in both the horizontal and the vertical directions. The objects on the top-most shelves were placed at a height of 190cm and on the bottom-most shelves at a height of 13cm from the ground level.

5.3 Experimental Task

Subjects performed two practice trials where they familiarized themselves with handling the VR controller and the experimental setup. In these practice trials, they were free to explore the virtual environment and displace the objects. After the practice trials, subjects were asked to sort objects based on the one and/or two features of the object. Each subject performed 24 trials in total, with each trial instruction (as listed below) randomly presented twice throughout the experiment. The experimental setup is illustrated in Figure 1A. The trial instructions were as follows:

1. Sort objects so that each row has the same shape or is empty
2. Sort objects so that each row has all unique shapes or is empty
3. Sort objects so that each row has the same color or is empty
4. Sort objects so that each row has all unique colors or is empty
5. Sort objects so that each column has the same shape or is empty
6. Sort objects so that each column has all unique shapes or is empty
7. Sort objects so that each column has the same color or is empty
8. Sort objects so that each column has all unique colors or is empty
9. Sort objects so that each row has all the unique colors and all the unique shapes once
10. Sort objects so that each column has all the unique colors and all the unique shapes once
11. Sort objects so that each row and column has each of the four colors once.
12. Sort objects so that each row and column has each of the four shapes once.

5.4 Data pre-processing

We measured the position and direction 3D vectors of the head and hand in global reference frame. The eye direction and position vectors were measured in the head reference frame. **Figure 1B** illustrates the vector representations of the eye, head, and hand position and orientation vectors. At the outset, we downsampled the data to 40 Hz. The sections below explain the steps we took to process the raw data and arrive at the magnitude of translation and rotation movements made by the eye, head, and hand.

Gaze data

Using the eye-in-head 3D gaze direction vector for the cyclopean eye, we calculated the gaze angles for the horizontal θ_{h} and vertical θ_{v} directions. vector samples were sorted by their timestamps. The 3D gaze direction vector of each sample is represented in (x, y, z) coordinates as a unit vector that defines the direction of the gaze in VR world space coordinates. In VR, the x coordinate corresponds to the left-right direction, y in the up-down direction, z in the forward-backward direction. From the eye-in-head gaze direction vectors, we computed the horizontal (θ_{h}) and the vertical (θ_{v}) gaze angles in degrees using the following formula:

$$\theta_{\mathrm{h}} = \frac{180}{\pi} \cdot \arctan \frac{x}{z} \tag{1}$$

$$\theta_{\mathrm{v}} = \frac{180}{\pi} \cdot \arctan \frac{y}{z} \tag{2}$$

HMD data

From the HMD we obtained the global head position vector and the head direction vector. Using the 3D head direction vector, angular orientation in the horizontal θ_{h} and vertical θ_{v} directions using the following formula:

$$\theta_{\mathrm{h}} = \frac{180}{\pi} \cdot \arctan \frac{x}{z} \tag{3}$$

$$\theta_{\mathrm{v}} = \frac{180}{\pi} \cdot \arctan \frac{y}{z} \tag{4}$$

At the beginning of each trial, subjects were asked to stand still facing the shelf for 3s at a set location. Thus, we could estimate the initial position of the head from the average position of the HMD in the 3s period. Subsequently, we calculated the deviations of the HMD from the initial position arrive at the magnitude and direction of translation movements. For each time point after the initial 3s period, we subtracted the initial position from the current position in 3D coordinates. Hence, we could estimate the degree of head translation movements in the left-right and up-down directions.

Hand controller data

Subjects used the trigger button of the HTC Vive controller to virtually grasp the objects on the shelf and displace them to other locations. In the data, the trigger was recorded as a boolean, which was set to TRUE when subjects pressed the trigger button on the hand controller to initiate an object displacement and was reset to FALSE when the trigger button was released, and the object was placed on the shelf. Using the position of the controller in the world space, we determined the locations from the shelf where a grasp was initiated and ended. We also removed trials where the controller data showed implausible locations in the 3D space. These faulty data were attributed to the loss of tracking during the experiment. Next, we removed grasping periods where the origin and final locations of the objects were the same on the shelf.

Next, we calculated the angular position of the hand with respect to the head positions in each trial. As described above, the x-coordinate corresponds to the left-right direction, y in the up-down direction, z in the forward-backward direction. Using the 3D Cartesian coordinates (x, y, z) of the controller position $(Hand_{(x,y,z)})$, HMD position $(Head_{(x,y,z)})$ and HMD direction unit vector $(\vec{\mathrm{Head}}_{(x,y,z)})$ in world space, we calculated the horizontal and the vertical angular position of the hand with respect to the head. The horizontal (ϕ_{h}) and vertical (ϕ_{h}) angular position of the hand was calculated as follows:

$$\phi_{\mathrm{h}} = \frac{180}{\pi} \cdot \arccos \left(\frac{Hand_{(x,z)} - Head_{(x,z)}}{\|Hand_{(x,z)} - Head_{(x,z)}\|} \cdot \vec{\mathrm{Head}}_{(x,z)} \right) \tag{5}$$

$$\phi_{\mathrm{v}} = \frac{180}{\pi} \cdot \arccos \left(\frac{Hand_{(y,z)} - Head_{(y,z)}}{\|Hand_{(y,z)} - Head_{(y,z)}\|} \cdot \vec{\mathrm{Head}}_{(y,z)} \right) \tag{6}$$

where $\|\|$ denotes the norm of the vector, and $(\cdot)$ denotes the dot product.

5.5 Data Analysis

After pre-processing, we were left with data from 27 subjects with 554 trials in total and $Mean = 20.51, SD = \pm 2.20$ trials per subject. Furthermore, we had eye-tracking data corresponding to 5664 grasping actions with $12.85, SD = \pm 1.91$ object displacements per trial and subject.

5.5.1 Principal Components Analysis

Given the naturalistic setting of our experimental setup, with complex movements of the head, eye, and hand, we wanted to understand the contributions and coupling of each of these effectors while performing object pickup and dropoff actions. We first epoched the data using the grasp onset and offset triggers. We selected a time window of 1s before and after the trigger. Thus, each epoch consisted of 3D data from head position, head rotation, eye direction, and hand position at each time point spaced 0.025s apart. Each subject's feature matrix then consisted of a 3D matrix of 80 time-points, each with 12 features, for each grasp. We explored the time-wise low-dimensional representation of this matrix using Principal Component Analysis.

For each subject, the input matrix for PCA was composed of $G \times M \times T$, where G denotes the grasping epoch, M denotes the 12 features (four data streams in x, y, z coordinates) and T denotes each time point relative to the action events. For each time point in T, we standardized $G \times M$ matrix to zero mean and unit standard deviation. This was done to make sure each feature contributed equally to the PCA. We then applied PCA to the $G \times M$ matrix at each time point. We used the 'sklearn.preprocessing.PCA' python package to compute the eigenvalues and principal components. We then used the 'transform()' function to reduce the dimensionality of the matrix from $G \times M$ to $G \times 2$. Thus, for each subject, we obtained $T \times 2$ eigenvalues and $2 \times M \times T$ coefficients of the eigenvectors. The eigenvalues were used to show the explained variance of the 2 principal components (PCs) across time. The coefficients provided the loadings of each M feature onto the PCs. Hence, for each subject, we could ascertain the evolution of the contribution of the features in the PCA subspace across time.

5.5.2 Cosine Similarity

To understand the contributions of the individual features to the PCA subspace, we calculated the directional similarity of the variables. For each eye-head, eye-hand, and head-hand pair, we computed the cosine similarity of their horizontal and vertical components as follows:

$$\cos\theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \cdot \|\vec{b}\|} \tag{7}$$

Using the values of $\cos\theta$ we could determine if the coefficients of the eigenvector were aligned in the same direction ($\cos\theta = 1$), opposite directions ($\cos\theta = -1$) or orthogonal ($\cos\theta = 0$) across the different time points.

5.5.3 Genralization & Predictions

To explore the generalizability of the PCA subspace across subjects. We used N-fold cross-validation, where N corresponds to the number of subjects. For each fold, we divided the dataset into train and test sets, where data from $N - 1$ subjects was used for training, and the left-out dataset from 1 subject was used for testing. During training, we applied PCA to $A \times M$ matrix at each time point of the grasping epochs, where A denotes all grasping epochs across $N - 1$ subjects. After applying PCA, we obtained the $A \times 2$ reduced matrix in PCA subspace. Using the $A \times 2$ matrix, we trained a kernel-based support vector machine (SVM) to predict the location of the action at each time point. We used the sklearn.svm.SVC python package for training

and prediction. We used the default parameters for the model and did not perform any hyper-parameter optimization. During testing, we standardized the $G \times M$ matrix of the left-out subject, applied the PCA weights from the training set to reduce its dimensionality to $G \times 2$, and recorded the mean prediction accuracy of the trained model on the left-out data. This analysis was repeated for the grasp offset events. In this manner, we could test the generalizability and predictive power of the PCA subspace across time for grasp onset and offset events.

Author Contributions

AK, PK, TS: conceived and designed the study. TS, PK: Procurement of funding. AK: data collection. AK, FB: data pre-processing. AK: data analysis. AK, MAW: initial draft of the manuscript. AK, MAW, TS, PK: revision and finalizing the manuscript. All authors contributed to the article and approved the submitted version.

We would like to thank Shadi Derakhshan and Imke Mayer for helping with the data collection.

Data availability statement

The experimental data and analysis code can be found at `https://osf.io/9edby/`.

Acknowledgement

We are grateful for the financial support by the German Federal Ministry of Education and Research for the project ErgoVR (Entwicklung eines Ergonomie-Analyse-Tools in der virtuellen Realität zur Planung von Arbeitsplätzen in der industriellen Fertigung)-16SV8052. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

H. K. Arora, V. Bharmauria, X. Yan, S. Sun, H. Wang, and J. D. Crawford. Eye-head-hand coordination during visually guided reaches in head-unrestrained macaques. *Journal of neurophysiology*, 122(5):1946–1961, Nov. 2019. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00072.2019. URL `https://www.physiology.org/doi/10.1152/jn.00072.2019`.

A. Belardinelli, M. Y. Stepper, and M. V. Butz. It's in the eyes: Planning precise manual actions before execution. *Journal of vision*, 16(1):18, 2016. ISSN 1534-7362. doi: 10.1167/16.1.18. URL `http://dx.doi.org/10.1167/16.1.18`.

W. Bialek. On the dimensionality of behavior. *Proceedings of the National Academy of Sciences*, 119(18): e2021860119, 2022. doi: 10.1073/pnas.2021860119. URL `https://www.pnas.org/doi/abs/10.1073/pnas.2021860119`.

B. Biguer, M. Jeannerod, and C. Prablanc. The role of position of gaze in movement accuracy. *Attention and performance XI*, pages 407–424, 1985.

M. C. Bowman, R. S. Johansson, and J. R. Flanagan. Eye-hand coordination in a sequential target contact task. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 195(2):273–283, May 2009. ISSN 0014-4819, 1432-1106. doi: 10.1007/s00221-009-1781-x. URL `http://link.springer.com/10.1007/s00221-009-1781-x`.

C. A. Buneo, M. R. Jarvis, A. P. Batista, and R. A. Andersen. Direct visuomotor transformations for reaching. *Nature*, 416(6881):632–636, Apr. 2002. ISSN 0028-0836. doi: 10.1038/416632a. URL http://dx.doi.org/10.1038/416632a.

J. C. Culham, C. Cavina-Pratesi, and A. Singhal. The role of parietal cortex in visuomotor control: what have we learned from neuroimaging? *Neuropsychologia*, 44(13):2668–2684, 2006. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2005.11.003. URL http://dx.doi.org/10.1016/j.neuropsychologia.2005.11.003.

F. R. Danion, J. Mathew, N. Gouirand, and E. Brenner. More precise tracking of horizontal than vertical target motion with both the eyes and hand. *Cortex; a journal devoted to the study of the nervous system and behavior*, 134:30–42, Jan. 2021. ISSN 0010-9452, 1973-8102. doi: 10.1016/j.cortex.2020.10.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0010945220303713.

J. A. Droll and M. M. Hayhoe. Trade-offs between gaze and working memory use. *Journal of experimental psychology. Human perception and performance*, 33(6):1352–1365, Dec. 2007. ISSN 0096-1523. doi: 10.1037/0096-1523.33.6.1352. URL http://dx.doi.org/10.1037/0096-1523.33.6.1352.

B. V. Ehinger, K. Groß, I. Ibs, and P. König. A new comprehensive eye-tracking test battery concurrently evaluating the pupil labs glasses and the EyeLink 1000. *PeerJ*, 7:e7086, July 2019. doi: 10.7717/peerj.7086. URL http://dx.doi.org/10.7717/peerj.7086.

J. R. Flanagan, M. C. Bowman, and R. S. Johansson. Control strategies in object manipulation tasks. *Current opinion in neurobiology*, 16(6):650–659, Dec. 2006. ISSN 0959-4388. doi: 10.1016/j.conb.2006.10.005. URL http://dx.doi.org/10.1016/j.conb.2006.10.005.

N. Fujii, H. Mushiake, and J. Tanji. Rostrocaudal distinction of the dorsal premotor area based on oculomotor involvement. *Journal of neurophysiology*, 83(3):1764–1769, Mar. 2000. ISSN 0022-3077. doi: 10.1152/jn.2000.83.3.1764. URL http://dx.doi.org/10.1152/jn.2000.83.3.1764.

K. Hadjidimitrakis. Coupling of head and hand movements during eye-head-hand coordination: there is more to reaching than meets eye. *Journal of neurophysiology*, 123(5):1579–1582, May 2020. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00099.2020. URL http://dx.doi.org/10.1152/jn.00099.2020.

M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz. Visual memory and motor planning in a natural task. *Journal of vision*, 3(1):49–63, 2003. ISSN 1534-7362. doi: 10.1167/3.1.6. URL http://dx.doi.org/10.1167/3.1.6.

Y. Hu and M. A. Goodale. Grasping after a delay shifts size-scaling from absolute to relative metrics. *Journal of cognitive neuroscience*, 12(5):856–868, Sept. 2000. ISSN 0898-929X. doi: 10.1162/089892900562462. URL http://dx.doi.org/10.1162/089892900562462.

J. N. Ingram and D. M. Wolpert. Naturalistic approaches to sensorimotor control. *Progress in brain research*, 191:3–29, 2011. ISSN 0079-6123, 1875-7855. doi: 10.1016/B978-0-444-53752-2.00016-3. URL http://dx.doi.org/10.1016/B978-0-444-53752-2.00016-3.

R. S. Johansson, G. Westling, A. Bäckström, and J. R. Flanagan. Eye-hand coordination in object manipulation. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 21(17):6917–6932, Sept. 2001. ISSN 0270-6474, 1529-2401. doi: 10.1523/jneurosci.21-17-06917.2001. URL https://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.21-17-06917.2001.

A. Keshava, A. Aumeistere, K. Izdebski, and P. Konig. Decoding task from oculomotor behavior in virtual reality. In *ACM Symposium on Eye Tracking Research and Applications*, number Article 30 in ETRA '20 Short Papers, pages 1–5, New York, NY, USA, June 2020. Association for Computing Machinery. ISBN 9781450371346. doi: 10.1145/3379156.3391338. URL https://doi.org/10.1145/3379156.3391338.

A. Keshava, N. Gottschewsky, S. Balle, F. N. Nezami, T. Schüler, and P. König. Action affordance affects proximal and distal goal-oriented planning. *The European journal of neuroscience*, 57(9):1546–1560, May 2023. ISSN 0953-816X, 1460-9568. doi: 10.1111/ejn.15963. URL http://dx.doi.org/10.1111/ejn.15963.

A. Keshava, F. N. Nezami, H. Neumann, K. Izdebski, T. Schüler, and P. König. Just-in-time: gaze guidance in natural behavior. Jan. 2024. URL https://www.biorxiv.org/content/10.1101/2021.01.29.428782v4.

S. U. König, A. Keshava, V. Clay, K. Rittershofer, N. Kuske, and P. König. Embodied spatial knowledge acquisition in immersive virtual reality: Comparison to map exploration. *Frontiers in Virtual Reality*, 2, 2021. ISSN 2673-4192. doi: 10.3389/frvir.2021.625548. URL https://www.frontiersin.org/articles/10.3389/frvir.2021.625548.

M. Land, N. Mennie, and J. Rusted. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311–1328, 1999. ISSN 0301-0066. doi: 10.1068/p2935. URL http://dx.doi.org/10.1068/p2935.

M. F. Land. Predictable eye-head coordination during driving. *Nature*, 359(6393):318–320, Sept. 1992. ISSN 0028-0836. doi: 10.1038/359318a0. URL http://dx.doi.org/10.1038/359318a0.

M. F. Land and S. Furneaux. The knowledge base of the oculomotor system. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 352(1358):1231–1239, Aug. 1997. ISSN 0962-8436. doi: 10.1098/rstb.1997.0105. URL http://dx.doi.org/10.1098/rstb.1997.0105.

M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision research*, 41(25-26):3559–3565, 2001. ISSN 0042-6989. doi: 10.1016/s0042-6989(01)00102-x. URL http://dx.doi.org/10.1016/s0042-6989(01)00102-x.

D. C. Niehorster, T. Santini, R. S. Hessels, I. T. C. Hooge, E. Kasneci, and M. Nyström. The impact of slippage on the data quality of head-worn eye trackers. *Behavior research methods*, 52(3):1140–1160, June 2020. ISSN 1554-351X, 1554-3528. doi: 10.3758/s13428-019-01307-0. URL http://dx.doi.org/10.3758/s13428-019-01307-0.

J. Pelz, M. Hayhoe, and R. Loeber. The coordination of eye, head, and hand movements in a natural task. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 139(3):266–277, Aug. 2001. ISSN 0014-4819. doi: 10.1007/s002210100745. URL http://dx.doi.org/10.1007/s002210100745.

T. D. Sanger. Human arm movements described by a low-dimensional superposition of principal components. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 20(3):1066–1072, Feb. 2000. ISSN 0270-6474, 1529-2401. URL https://www.ncbi.nlm.nih.gov/pubmed/10648712.

M. Santello, M. Flanders, and J. F. Soechting. Postural hand synergies for tool use. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 18(23):10105–10115, Dec. 1998. ISSN 0270-6474. doi: 10.1523/JNEUROSCI.18-23-10105.1998. URL http://dx.doi.org/10.1523/JNEUROSCI.18-23-10105.1998.

J. B. Smeets, M. M. Hayhoe, and D. H. Ballard. Goal-directed arm movements change eye-head coordination. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 109(3):434–440, June 1996. ISSN 0014-4819. doi: 10.1007/BF00229627. URL http://dx.doi.org/10.1007/BF00229627.

A. Stamenkovic, P. J. Stapley, R. Robins, and M. A. Hollands. Do postural constraints affect eye, head, and arm coordination? *Journal of neurophysiology*, 120(4):2066–2082, Oct. 2018. ISSN 0022-3077, 1522-1598. doi: 10.1152/jn.00200.2018. URL http://dx.doi.org/10.1152/jn.00200.2018.

T. A. Tu and E. G. Keating. Electrical stimulation of the frontal eye field in a monkey produces combined eye and head movements. *Journal of neurophysiology*, 84(2):1103–1106, Aug. 2000. ISSN 0022-3077. doi: 10.1152/jn.2000.84.2.1103. URL http://dx.doi.org/10.1152/jn.2000.84.2.1103.