Detection of task-relevant and task-irrelevant motion sequences: application to motor adaptation in goal-directed and whole-body movements

Daisuke Furuki[1], **Ken Takiyama**[1]
[1]Department of Electrical and Electronic Engineering, Tokyo University of Agriculture and Technology, Koganei-shi, Tokyo 184-8588, Japan, Correspondence should be addressed to K.T. (t.j.ken.takiyama@gmail.com).

Abstract

Motor variability is inevitable in our body movements and has been discussed from various perspectives in motor neuroscience and biomechanics; it can originate from the variability of neural activities, reflect a large degree of freedom inherent in our body movements, decrease muscle fatigue, and facilitate motor learning. How to evaluate this motor variability is thus a fundamental question in motor neuroscience and biomechanics. Previous methods quantified (at least) two striking features of motor variability: a smaller variability in the task-relevant dimension than in the task-irrelevant dimension and a low-dimensional structure that have often been referred to as synergy or the principal component. However, these previous methods were not only unsuitable for quantifying these features simultaneously but also applicable in only limited conditions (e.g., one method cannot consider motion sequence and another method cannot consider how each motion is relevant to performance). Here, we propose a flexible and straightforward machine learning technique that can quantify task-relevant variability, task-irrelevant variability, and the relevance of each principal component to task performance while considering motion sequence and the relevance of each motion sequence to task performance in a data-driven manner. We validate our method by constructing a novel experimental setting to investigate goal-directed and whole-body movements. Further, our setting enables us to induce motor adaptation by using perturbation and evaluate the modulation of task-relevant and task-irrelevant variabilities through motor adaptation. Our method enables to identify a novel property of motor variability: the modulation of those variabilities is different depending on perturbation schedule. A constant perturbation increases task-relevant variability, and a gradually imposed perturbation increases task-irrelevant variability.

Introduction

In our daily life, we can repeatedly achieve desired movements, such as grasping a cup, throwing a ball, and playing piano. To achieve the desired movements, our motor system needs to resolve at least two difficulties inherent in our body motion [1]. One difficulty is movement variability. Due to the variability inherent in various stages such as acquiring sensory information [2], the neural activities of motor planning [3], or the muscle activities during motor execution [4], even sophisticated athletes and musicians can not repeat a given movement. Our motor systems somehow overcome these variabilities to achieve the desired

movements [5]. Another difficulty is a large degree of freedom (DoF) inherent in our motor system [1, 6]. The number of joints, muscles, and neurons are more than necessary to achieve the desired movements, resulting in that an infinite number of joint configurations, muscle activities, and neural activities that can result in the desired movement [7-10]. Our motor systems somehow resolve these difficulties (i.e., variability and a large DoF) and generate the desired movements.

Although it remains unclear how we overcome movement variability, a possible answer is the decomposition of motor variability into task-relevant and task-irrelevant variabilities. We compensate for a portion of the motor variability that is relevant to achieve desired movements (i.e., task-relevant variability) [11-15]; simultaneously, we do not significantly compensate for a portion of the variability that is irrelevant to achieve the desired movements (i.e., task-irrelevant variability). The compensation of task-relevant variability can be observed in movement kinematics [11-14], muscle activities [16, 17], and neural activities [15]. This striking feature of our motor variability enables us to achieve a desired movement under the existence of movement variability.

Several studies have developed the techniques to evaluate task-relevant and task-irrelevant variabilities. The uncontrolled manifold (UCM) evaluates the task-relevant and task-irrelevant variabilities (mainly) in joint angles and angular velocities. The method focuses on the kinematic parameters relevant to task achievement, such as hip joint position in stand-and-sit motion [11] or hand position in arm-reaching movements [18]. The Jacobian matrix, the derivatives of the kinematic parameters concerning joint angles or angular velocities, enables us to define null space around the joint angles or angular velocities averaged across trials. The variability along the null space can be defined as task-irrelevant variability. Previous studies have revealed that the task-relevant variability of joint angles and angular velocities is less than the task-irrelevant variability. Notably, the UCM focuses on forward kinematics that map joint angles and angular velocities onto joint positions and velocities in the external coordinate. In contrast, tolerance, noise, and covariation analysis (TNC) [13] and goal-equivalent manifold analysis (GEM) [14] focus on task functions that define the relationship between kinematic parameters and task performance. For example, a thrown dart or ball can be modeled as a parabola. When the release position and velocity in the vertical axis are p and v, respectively, the maximum height of the released dart or ball can be written as $h = p + \frac{v^2}{2g}$, where g is gravitational acceleration. When controlling h is a task, the relation among h, p, and v can be task function. For example, with a slight value d (i.e., $d^2 \simeq 0$), a slight change in the release position $p + \frac{v}{g}d$ and release velocity $v - d$ does not cause any change in h, allowing the variability along these slight changes to be regarded as task-irrelevant variability. TNC and GEM evaluate the task-relevant and task-irrelevant variabilities based on the task functions.

These techniques have pros and cons. The UCM enables us to evaluate motion sequence, but it does not consider task function. The framework is thus not always suitable for the situation in which kinematic parameters are nonlinearly relevant to task achievements, as is the case with the quadratic function of v in the parabolic mentioned above. Because forward kinematics are nonlinear functions of joint angles and angular velocities, the UCM requires local linear approximation around representative joint angles or angular velocities based on the Jacobian matrix. Due to linear approximation, the UCM assumes the variability around the kinematics averaged across all trials. This approximation results in difficulty considering the motor variability when the averaged kinematics change, as with the kinematics before, during, and after motor learning (although it is possible to discuss these situations separately [18]). GEM also considers the local linear approximation of nonlinear task functions, resulting in the method considering variability around the task parameters averaged across all trials. Although GEM can deal with task function, it is difficult to apply to motion sequence in several cases. For example, to consider a motion sequence, the framework needs to define how the dart or ball position and velocity 100 msec before release are relevant to maximum height in the example mentioned above. TNC enables simultaneous discussion of the motor variability before, during, and after motor learning because the framework captures the entire variability in a nonparametric manner without local linear approximation; however, TNC is not always suitable for considering motion sequence based for reasons similar to those facing GEM: it requires explicit

definition of task function. In total, each method has pros and cons, resulting in no single framework that can simultaneously evaluate task-relevant and task-irrelevant variabilities when the averaged kinematics or task parameters change while considering both motion sequence and task function.

Motor variability has another striking feature: the variability is embedded in low-dimensional space that is referred to as synergy [6, 19-22]. It is suggested that to overcome a larger DoF, our motor system controls not all the DoFs but only those in low-dimensional space. Synergy has been (mainly) discussed for kinematics data [20, 22-24] and EMG data [19, 21]. In both kinematics and EMG, low-dimensional space that can capture a high portion of motor variability has been found. Several methods have been developed to extract the synergy, such as principal component analysis (PCA) [20, 22], nonnegative matrix factorization [21, 25] or spatial-temporal decomposition of EMG data [19].

Motor variability thus has at least two characteristics: compensation of task-relevant variability and low-dimensional structure. Most of the techniques, however, deal with only one of these aspects. It is difficult to detect the low-dimensional structure of motor variability by the methods that evaluate task-relevant and task-irrelevant variabilities. Similarly, it is difficult to evaluate task-relevant and task-irrelevant variabilities by the techniques that extract the low-dimensional structure of motor variability. In other words, the 1st principal component, the dimension that can explain the most significant portion of variability among all dimensions, is not always the factor most relevant or irrelevant to task performance. Although the UCM had been used to extract synergy [26], the primary advantage of the UCM is not the extraction of low-dimensional structure but the evaluation of task-relevant and task-irrelevant variabilities. Although a few studies have focused on linear discrimination analysis (LDA) to discuss task-relevant low-dimensional space [27, 28], LDA enables only discrimination, i.e., success or failure of the movement [29], in contrast to TNC and GEM, which can account for motor performance based on continuous performance values. In summary, few methods simultaneously quantify the two striking features of movement variability.

Here, we propose a flexible and straightforward machine learning technique that can evaluate movement variability by unifying the advantages of previous techniques; our framework can evaluate not only task-relevant and task-irrelevant variabilities even when the averaged kinematics or task parameters change (e.g., before, during, and after motor learning) while considering motion sequence and task function but also how each synergy is relevant to task performance by extending the PCA. The current study relied on ridge regression [30], a linear regression technique that is robust to measurement noise, has a definite relation to the PCA and can evaluate how the motion of each body part at each time point is relevant to task performance in a data-driven manner [31]. Our technique can thus enable us to determine the task functions in a data-driven manner without any explicit function such as the parabola. First, we construct a novel experimental paradigm to discuss the relation of a motion sequence to task performance based on goal-directed and whole-body movements. We further discuss motor adaptation in the current experimental setting. Second, we formalize the dissociation of motion sequence data into task-relevant components and task-irrelevant components by extending a ridge regression. We validate the decomposition based on our experimental data. Third, we clarify the relation between ridge regression and the PCA, a popular method to extract motor variability in low-dimensional space. In particular, we analytically reveal how each principal component is relevant to performance in the ridge regression. We also validate the analytical calculations based on our experimental data. Finally, we apply our method to motion sequence data in whole-body and goal-directed movements before and after motor adaptation. Because our method enables us to discuss the modulation of movement variability before and after motor adaptation, we discuss the dependence of the modulation on perturbation schedule.

Results

Goal-directed whole-body movements and motor adaptation

The current study focuses on goal-directed and whole-body movements in which subjects manage to achieve a desired movement by controlling a large DoFs. We focus on a simplified version of whole-body movements: a vertical jump while crossing arms in front of the trunk (Fig. 1A). This goal-directed whole-body movement enabled us to focus on lower limb motions and assess task-relevant variability, task-irrelevant variability, and the low-dimensional spaces in which a high portion of the motor variability was embedded. We proposed a machine learning technique to simultaneously evaluate these features of variability while considering motion sequence and the relevance of each motion to jumping height.

Subjects stood in a fixed position and were instructed to look at a computer monitor located in front of them and perform a sub-maximum vertical jump according to a target height (40, 45, 50, 55, or 60% of the maximum jumping height of each subject, Fig. 1B). Three beeps sounded, and the subjects needed to perform the jump at the timing of the third beep. The interval between each beep was one second. At the beginning of each trial (i.e., one second before the first beep), the target height was indicated by a black bar displayed on a computer monitor. At the end of the tth trial, the actual jumping height h_t (the y-position of the marker attached to the back) was displayed as a blue cursor on the monitor, where $t = 1, ..., T$ and T was the number of trials to be analyzed. By manipulating the displayed jumping height (we called this manipulation as a perturbation p_t), it was possible to induce sensory prediction error between the predicted and actual jumping heights. This perturbation paradigm was similar to a protocol of motor gain adaptation as reported mainly in saccade and arm-reaching movements [32, 33]. We expected subjects to modify their motion sequences to minimize the sensory prediction error.

First, we determined whether the subjects could perform goal-directed whole-body movements in our experimental setting. In 50 baseline trials in experiment 1 (Fig. 1C), the target height pseudorandomly changed in each trial. There were significant differences in jumping height depending on target height (Fig. 1D, 1-way ANOVA, p = 2.2597×10^{-21}). The subjects could thus perform goal-directed vertical jump depending on target height.

Second, we determined whether the subjects showed motor adaptation in the experimental setting. In 96 learning trials in experiment 1 (Fig. 1C), the subjects experienced perturbations once in every five trials; the perturbation was pseudorandomly set to $p_t = 0.05$ or $p_t = -0.05$ in every five trials and $p_t = 0$ in other trials (Figs. 2A and B). We observed the modification of jumping height after each perturbation (Fig. 2C, paired t-test, $p = 0.0026$ for motor adaptation when $p_t = 0.05$; and $p = 0.0014$ for motor adaptation when $p_t = -0.05$). Motor adaptation could thus be observed in the goal-directed vertical jump.

Validation of ridge regression and decomposition into task-relevant and task-irrelevant components

The current study focuses on the evaluation of motor variability, especially task-relevant variability, task-irrelevant variability, and the relevance of low-dimensional structures to task performance, by extending ridge regression (the details of ridge regression were provided in Methods). Ridge regression is a linear regression technique that is robust against observation noise, is applicable to data with multicollinearity, and enables predictions of output data y (i.e., jumping height) based on input data $\boldsymbol{X} = (X_1, ..., X_D)$ (i.e., motion sequence). The regression thus enables the prediction of jumping height as $h = \boldsymbol{X}\boldsymbol{w} = \sum_{i=1}^{D} X_i w_i$, where h is predicted jumping height and $\boldsymbol{w} = (w_1, ..., w_D)^T$ is weight coefficients to predict jumping height, where $(a_1, ...)^T$ denotes the transpose of the vector. When the absolute value of w_i is a large value, X_i can be considered to be relevant to task performance. In contrast, when w_i is close to 0, X_i can be considered to be irrelevant to task performance. We needed to validate the ridge regression in the current experimental setting before evaluating variability. Notably, we have already validated the

efficiency of ridge regression to predict performance not only in jumping movements but also in throwing movements [31].

Ridge regression requires selecting input data because a careful selection of input data is indispensable to discussing the linear relation between input and output data. Prediction power is a sophisticated measure for selecting input data while avoiding overfitting [29]. The current study focuses on prediction error between actual and predicted jumping height using 10-fold cross validation. We compared the following three types of input data (see Methods for details). The first candidate is joint angles $\{q_i\}$ and angular velocities $\{\dot{q}_i\}$: $\boldsymbol{X} = (\{q_i\}, \{\dot{q}_i\})$, where $\{a_i\} = (a_1, a_2, a_3)$ and $\dot{q}_i$ denotes the derivative of q_i concerning time (the definitions of each q_i are given in Fig. 1A). The second candidate is the functions of q_i and $\dot{q}_i$, which describe the position and velocity of hip joint in the y-axis and are relevant to jumping height: $\boldsymbol{X} = (\{\sin q_i\}, \{\dot{q}_i \cos q_i\})$. The third candidate is the functions describing the jumping height based on parabolic approximation: $\boldsymbol{X} = (\{\sin q_i\}, \{\dot{q}_i \dot{q}_j \cos q_i \cos q_j\})$, where $\{a_i a_j\} = (a_1^2, a_1 a_2, a_1 a_3, a_2^2, a_2 a_3, a_3^2)$. By comparing these three candidates, we found that the third candidate showed the lowest prediction error (Fig. 3A). In particular, the third candidate, with seven time frames before release, yielded the lowest prediction error. In the following, we refer to the third candidate with seven time frames as the motion sequence. If the prediction error equals 1, the method cannot predict the output data. In contrast, if the prediction error equals 0, the method can predict the output data with 100% accuracy. As shown in Fig. 3A, the third candidate with seven time frames resulted in a prediction error of less than 0.2, indicating that the ridge regression enables prediction of jumping height with a greater than 80% accuracy in the current setting. Because the parabolic approximation of jumping height ($h = p + \frac{v^2}{2g}$, the detailed definition is given in the Introduction) enables prediction of the jumping height with a 76% accuracy (purple line in Fig. 3A), the ridge regression enables prediction of jumping height with higher accuracy than the approximation. The reasons the ridge regression shows higher prediction power are its robustness against observation noise and consideration of the motion sequence rather than the representative motion data at a single time frame (i.e., the position and velocity of the hip joint only at the time of release). The ridge regression thus enables the discussion of the linear relation between the motion sequence and jumping height with appropriate precision.

Based on the ridge regression, the current study decomposes the motion sequence $\boldsymbol{X}$ into a task-relevant $\boldsymbol{X}_{\mathrm{rel}} = \boldsymbol{X} \frac{\boldsymbol{w}\boldsymbol{w}^T}{|\boldsymbol{w}|^2}$, where $|\boldsymbol{w}| = \sqrt{\boldsymbol{w}^T \boldsymbol{w}}$ is the squared norm of $\boldsymbol{w}$, and a task-irrelevant components $\boldsymbol{X}_{\mathrm{irr}} = \boldsymbol{X} - \boldsymbol{X}_{\mathrm{rel}}$ (see Methods for details). Figs. 3B and 3C demonstrate typical examples of the decomposition when $\boldsymbol{X}$ includes only 2 elements and constrains the task by setting $y = X_1 - X_2$ (i.e., $w_1 = 1$ and $w_2 = -1$) to some certain values (e.g., y = 2, 0, -2 in the simulated task 1, 2, 3, respectively). Because the constrained task was one dimensional and input data were two dimensional, an infinite number of $\boldsymbol{X}$ values resulted in an identical y value. In this case, $\boldsymbol{X}_{\mathrm{rel}} = \boldsymbol{X} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{\sqrt{2}}(X_1 - X_2, -(X_1 - X_2))$ and $\boldsymbol{X}_{\mathrm{irr}} = \boldsymbol{X} - \boldsymbol{X}_{\mathrm{rel}}$. The simulated data points on the dotted line in Fig. 3C indicated $\boldsymbol{X}_{\mathrm{rel}}$. On the dotted line, the data points can be clearly separated into three parts corresponding to the simulated tasks 1, 2, and 3. In contrast, on the solid line, $\boldsymbol{X}_{\mathrm{irr}}$ is not separated based on task. In our experimental setting with goal-directed vertical jump, $\boldsymbol{X}$, $\boldsymbol{X}_{\mathrm{rel}}$, and $\boldsymbol{X}_{\mathrm{irr}}$ included 63 elements in each trials (3(dim)×7(time frames) for $\{\sin q_i\}$, and 6(dim)×7(time frames) for $\{\dot{q}_i \dot{q}_j \sin q_i \sin q_j\}$. The current study calculated the variability (variance) of each element of $\boldsymbol{X}_{\mathrm{rel}}$ and $\boldsymbol{X}_{\mathrm{irr}}$ in focused trials. The representative values of the variability, task-relevant variability $\mathrm{Var}_{\mathrm{rel}}$ and task-irrelevant variability $\mathrm{Var}_{\mathrm{irr}}$, were calculated by averaging the variability across all the elements.

Variability in task-relevant and task-irrelevant space

We calculated the task-relevant and task-irrelevant variabilities in a goal-directed vertical jump based on both the ridge regression and the decomposition of input data into task-relevant and task-irrelevant dimensions. We found that the task-relevant variability was smaller than the task-irrelevant variability

in all participants (N = 13, red dots in Fig. 4A). Because previous methods, such as the UCM (blue crosses) and GEM (green crosses), found similar task-relevant and task-irrelevant variabilities, our method enabled the extraction of lower task-relevant variability and higher task-irrelevant variability. Because the normalization procedures of our method and previous methods differ (see Methods), there was a slight difference in the calculated variabilities. Our methods can quantify task-relevant and task-irrelevant variabilities by considering the motion sequence and the relevance of the sequence to the task. Our method does not require any explicit task function, such as the parabolic approximation of jumping height, but it determines the relevance of the motion sequence to the task in a data-driven manner. Further, our method is robust against observation noise due to the ridge regression.

Relevance of each principal component to task performance

Movement variability shows not only less task-relevant variability than task-irrelevant variability but also a low-dimensional structure. The current study compares our method to PCA, a conventional method to extract low-dimensional structure. We decomposed the motion sequence $\boldsymbol{X}$ into principal components (PCs, i.e., eigenvectors) and calculated the correlation of each PC to jumping height (see Methods for detail). When the motion sequence was decomposed using a larger number of principal components, the decomposed sequence showed a higher correlation to jumping height (Fig. 4B). If averaged across all participants, the 1st PC could explain approximately 40% of the movement variability (blue line in Fig. 4C). Corresponding to the explained movement variability, the 1st PC showed the highest correlation to jumping height (red line in Fig. 4C) if averaged across all participants. In a typical subject, however, the 2nd rather than the 1st PC showed the highest correlation to jumping height (red line in Fig. 4D). This typical subject was not an exception; Fig. 4E shows the PC number with the highest correlation to jumping height. In 6 out of 13 subjects, the 1st PC showed the highest correlation to performance. In 5 out of 13 subjects, the 2nd PC showed the highest correlation, and the 3rd PC showed the highest in 2 out of 13 subjects. These results indicate that the explained movement variability did not correspond to the relevance to task performance.

Ridge regression enabled the prediction of jumping height with higher accuracy than PCA (red line in Fig. 4B) because the ridge regression weights each PC based on both the explained movement variability and the task relevance. In PCA (or equivalently singular value decomposition (SVD)), the motion sequence at the tth trial is decomposed as $\boldsymbol{X}_t = \sum_{i=1}^{N} \lambda_i u_{i,t} \boldsymbol{v}_i$, where N is the number of PCs, λ_i is the ith eigenvalue corresponding to the ith PC $\boldsymbol{v}_i$, and $u_{i,t}$ indicated how the ith PC appeared at the trial. The correlation of the ith PC to task performance was thus calculated based on $u_{i,t}$, and it did not reflect the relevance of the ith PC to task performance. In contrast, the ridge regression enables the prediction of task performance as $h_t = \sum_{i=1}^{N} f(\lambda_i)\mathrm{Corr}(u_{i,t}, y_t)u_{i,t}$, where $f(\lambda_i)$ is a sigmoidal function of λ_i and $\mathrm{Corr}(u_{i,t}, y_t)$ is the correlation between the contribution of the ith PC at the tth trial $u_{i,t}$ and observed jumping height y_t (see Methods for details). The ridge regression thus enables the prediction of performance by weighting each PC based on both explained movement variability and task relevance. In other words, our method enables the consideration of the low-dimensional structure of movement variability by weighting each PC suitable for predicting task performance.

Influence of motor adaptation on variability in task-relevant and task-irrelevant dimension

An advantage of our method is its linearity, which enables the simultaneous comparison of the task-relevant and task-irrelevant variabilities among the conditions where mean kinematics or task parameters change (e.g., before, during, and after motor learning). It was previously unclear how task-relevant and task-irrelevant variabilities are modulated by motor adaptation. The modulation of these variabilities has been investigated for arm-reaching movements and motor adaptation to a constant perturbation [5, 34]. Although there are some differences between adaptation to a constant perturbation and that

to a gradually imposed perturbation, e.g., retention rate or awareness [35], the means by which those variabilities are modulated in the two types of adaptations have not been investigated. Further, it was previously unclear whether such modulation of variability could be observed in whole-body movements. Our method without linear approximation enabled the discussion of how task-relevant and task-irrelevant variabilities are modulated before and after motor adaptation in whole-body movements. We thus applied our method to motor adaptation in response to constant and gradually imposed perturbations.

In experiment 2 (two days for each subject), subjects experienced gradually increased or decreased perturbations. Each subject underwent ten learning trials without any perturbation. The perturbation gradually increased or decreased for ten trials and was set to 0.05 or -0.05 for ten trials (Figs. 5A, B). In a total of 30 trials, the target height was set to 50% of the subject' s maximum jumping height. Subjects who experienced a $p_t > 0$ on the first day experienced a $p_t < 0$ on the 2nd day and vice versa. The order of perturbation was counterbalanced across subjects. The subjects could adapt to the gradually increased or decreased perturbations (Fig. 5C).

In experiment 3, the subjects experienced constant perturbations. Each subject underwent five learning trials without any perturbation. The perturbation was set to 0.05 or -0.05 for 15 trials, 0 for ten trials for washout and -0.05 or 0.05 for 15 trials (Figs. 5D, E). Subjects who experienced a pt=0.05 on the 6th-20th trials experienced a $p_t = -0.05$ on the 31st-45th trials and vice versa. The order of perturbation was counterbalanced across subjects. In a total of 45 trials, the target height was set to 50% of the subject' s maximum jumping height. In both experiments 2 and 3, the subjects adapted to the perturbations (Fig. 5F).

We calculated the task-relevant and task-irrelevant variabilities before and after adaptation in experiments 2 and 3 (Figs. 6A and B). For task-relevant variability, there was no significant difference before and after the adaptation to gradually increasing or decreasing perturbations (blue dots in Fig. 6A, N = 13, Wilcoxon signed rank test, p = 0.1272). In contrast, in adapting to a constant perturbation, there was a significant difference in task-relevant variability before and after adaptation (red dots in Fig. 6A, N = 13, Wilcoxon signed rank test, p = 0.0105). For task-irrelevant variability, there was a significant difference before and after adaptation to gradually increasing or decreasing perturbations (blue dots in Fig. 6B, N = 13, Wilcoxon signed rank test, p = 0.0215). In contrast, in adapting to a constant perturbation, there was no significant difference in task-irrelevant variability before and after adaptation (red dots in Fig. 6B, N = 13, Wilcoxon signed rank test, p = 0.5417). These results could be interpreted based on a simple and two-dimensional case similar to that shown in Figs. 3B and C (Figs. 6C, D). In adapting to perturbations, the subjects needed to modify their output (i.e., jumping height) by determining an appropriate input (i.e., motion sequence). In adapting to gradually increasing or decreasing perturbations, the task-irrelevant variability increased, while the task-relevant variability was not modulated (Fig. 6C). In adapting to a constant perturbation, the task-relevant variability increased, while the task-irrelevant variability was not be modulated (Fig. 6D). In summary, the modulation of task-relevant and task-irrelevant variabilities depends on the schedule of perturbation.

Discussion

We proposed a flexible and straightforward machine learning technique that quantified task-relevant variability, task-irrelevant variability, and the relevance of each principal component to task performance in a noise-robust manner while considering motion sequence and how each motion sequence was relevant to task performance (Figs. 3 and 4). Our method can find the relevance of each motion sequence to performance (i.e., task function) in a data-driven manner; our method does not require any explicit task function, such as the parabolic approximation of jumping height. Further, our method does not require any linear approximation, which enables the simultaneous consideration of the variabilities when the kinematics or task parameters averaged across trials change (e.g., before, during, and after adaptation). By applying our method to the motion sequence before and after motor adaptation, we found that the

perturbation schedules affected the modulation of movement variability in motor adaptation (Figs. 6A and 6B). These advantages enable the methods to be flexibly applied to a wide range of goal-directed movements.

Another advantage of our method is the ability to select appropriate input based on predictive power (Fig. 3A). The predictive power also enables the selection of an appropriate coordinate to define task performance. Although we considered one-dimensional performance in the current study (i.e., jumping height), more than two-dimensional performance requires the definition of an appropriate coordinate [31, 36, 37]. Predictive power also plays an important role in selecting the appropriate coordinate [31].

Our method yields a state-space model for motor adaptation during whole-body movements. The state-space model was previously proposed as a model of motor adaptation mainly in arm-reaching movements [38-45]. In the current study, the modification of jumping height at the tth trial, $h_t - h_{t-1}$, was significantly correlated with the error, e_{t-1}, caused by perturbation and motor noise (in experiment 1, the correlation between $h_t - h_{t-1}$ and e_{t-1} averaged across all participants was 0.5118 and $p < 0.01$ for all the participants). The state-space model of jumping height can thus be written as $h_t = h_{t-1} + \eta e_{t-1}$, where $\eta (> 0)$ is the learning rate. The jumping height ht was predicted well by $h_t \simeq \boldsymbol{X}_t \boldsymbol{w} = \boldsymbol{X}_{\mathrm{rel,t}} \boldsymbol{w}$, which enabled us to approximately rewrite the state-space model as $\boldsymbol{X}_{\mathrm{rel,t}} \boldsymbol{w} = \boldsymbol{X}_{\mathrm{rel,t-1}} \boldsymbol{w} + \eta e_{t-1}$. The model indicated that the jumping height was modified via modifying the motion sequence in the dimension along $\boldsymbol{w}$.

We relied on a simple linear regression (i.e., ridge regression). It is possible to use a more complicated machine learning technique, such as a mixture model [29, 46-48], sparse regression technique [49], or non-linear regression technique [50]. We have shown that a nonlinear regression technique, such as Gaussian process regression, is not effective in predicting performance based on motion data [31], likely because the number of data is limited. Although sparse regression, nonlinear regression, or a mixture model can show better predictive performance if the number of the data is high enough in general, it is difficult to find certain relations between the principal components and estimated parameters via those methods. Ridge regression enables the determination of not only task-relevant and task-irrelevant variabilities but also the relevance of each PC to performance.

To our knowledge, only a few studies have investigated how variability is modulated through motor adaptation [5, 34]. A previous study clarified that the variability is modulated after motor adaptation by utilizing a constant force field [34]. To our knowledge, it was not clarified whether the perturbation schedule affected the modulation. The current study suggests that the perturbation schedule affects the modulation of variability (Figs. 6A and B). Because the variability can facilitate exploration [34], the current study also suggests that constant perturbations facilitate exploration in task-relevant space and that gradually applied perturbations affect the exploration less than constant perturbations. In contrast, a gradually applied perturbation can facilitate exploration in task-irrelevant space, and constant perturbations affect the exploration less than the gradually applied perturbations.

Materials and Methods

Participants. Thirteen healthy volunteers (aged 18-22 years, two females) participated in all of our experiments. On the first day, the participants underwent 10 practice trials and 160 baseline trials with pseudorandomly changing targets (40%, 45%, 50%, 55%, or 60% of the maximum jump height) and became accustomed to the experimental setting. At the second, third, fourth, and fifth days (not consecutive), they joined experiments 1, 2, and 3. They joined experiment 2 for two days. All participants were informed of the experimental procedures and their confirmation with the Declaration of Helsinki, and all participants provided written informed consent before the initiation of the experiments. All procedures were approved by the ethics committee of the Tokyo University of Agriculture and Technology.

Data acquisition and processing. Jumping motions were recorded at 120 Hz using six cameras (Optitrack Flex 13, NaturalPoint Inc., Corvallis, Oregon). Markers were attached to the back (TV10),

right hip joint (Femur greater Trochanter), right knee (Femur Lateral Epicondyle and Femur Medial Epicondyle), right foot (Fibula Apex of Lateral Malleolus and Tibia Apex of Medial Malleolus), and right toe (Head of 2nd Metatarsus) of the participants. Marker position data were filtered using a 12th-order, 10 Hz zero-phase Butterworth filter using MATLAB 2016a. Joint angles between the right toe and foot (q_1), right foot and shank (q_2), and right thigh and shank (q_3) were calculated in the sagittal plane (Fig. 1A). Because the current study focused on a vertical jump while crossing arms in front of the trunk, it was possible to focus only on lower limb motions. Throughout the current study, we focused on the three-link model of the lower limbs in the sagittal plane.

Release timing was detected based on the moment at which the vertical toe position exceeded 10% of the maximum height in each trial. The predictive power was calculated using various time bin lengths including the release timing (Fig. 3A). When the time bin length was seven, the seventh time frame corresponded to the release timing, the sixth time frame corresponded to one time frame before the release timing, and the other time frames followed accordingly.

Experimental setup. At the beginning of each trial, the subjects were instructed to stand at a fixed position. In each trial, subjects listened to three beeps separated by one-second intervals; the first beep indicated the start of each trial, and the subjects were required to jump at the timing of the third beep.

We measured the position of the marker attached to the subject' s back using MATLAB at 30 Hz. In front of the subject (1.5 meters ahead, 1.7 meters above the floor), there was a monitor to with a blue cursor that indicated the height of the marker attached to subject' s back and a black bar that indicated target height (Fig. 1B). Those cursors and bars were displayed one second before the first beep sounded. The blue marker could move only in the vertical axis because the current study focused on the vertical height of the jumping motion. The marker position at time t in the y-axis (Fig. 1A), y_t, was displayed on the monitor after being normalized for each subject as $y_t = \frac{\hat{y}_t - y_0}{y_{\max} - y_0}$, where $\hat{y}_t$, where $\hat{y}_t$ is the marker position without normalization, y_0 indicates the initial marker position at each trial, and $y_{\max}$ indicates the jumping height with maximum effort in each subject (Fig. 1C). For the two trials that required the subjects to jump with maximum effort, there was no cursor feedback. One second before the first beep, $y_t = 0$, the blue circle was displayed on the black baseline on the monitor. Additionally, the target height d was indicated by a black line. Before the baseline trials, the subjects underwent ten practice trials. In those trials, the marker position in each time frame was displayed on the monitor and d was pseudorandomly chosen from 0.40, 0.45, 0.50, 0.55, or 0.60 (each value was randomly chosen only once in every five trials). This method enabled the subjects to become acquainted with the experimental setting by confirming the motion trajectory of the marker attached to their back. In baseline trials, the marker position was displayed only at the start and end of each trial. One second before the first beep, the cursor was displayed on baseline position, and a black line was corresponding to d was displayed as pseudorandomly chosen from 0.40, 0.45, 0.50, 0.55, or 0.60. At the end of each trial, the cursor was displayed at the maximum value of y_t within each trial, which indicated jumping height (Fig. 1B). When the subjects achieved a jumping motion that was close to the target height ($|d - y_t| < 0.02$), they heard a coin-getting sound to indicate that the jumping motion was successful. After the baseline trials, the subjects underwent 96 learning trials in experiment 1, 30 trials in experiment 2 (the same set of practice and main trials was imposed for two days), and 45 trials in experiment 3.

We utilized a perturbation paradigm to investigate how subjects modify their jumping motion via experiencing sensory prediction errors. For trials with perturbation p_t, the position of the cursor was displayed at $y_t + p_t$. The subjects needed to modify their jumping motion to achiever a lower (when $p > 0$) or higher jumping height (when $p < 0$). When the displayed jumping height was close to the target height ($|d - (y_t + p_t)| < 0.02$), the subjects heard a coin-getting sound to indicate that the jumping motion was successful.

Decomposition into task-relevant and task-irrelevant components The current study relied on linear regression to determine the relation between input data $\boldsymbol{X} \in \mathrm{R}^{T \times D}$ and output data $\boldsymbol{d} \in \mathrm{R}^{T \times 1}$

following $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{w}$, where $\boldsymbol{y} \in \mathrm{R}^{T\times 1}$ is the predicted output data and $\boldsymbol{w} \in \mathrm{R}^{D\times 1}$ is the best linear coefficients to predict the output data. Although we relied on a ridge regression to estimate $\boldsymbol{w}$ (see below), the following decomposition of input data into the task-relevant and task-irrelevant components could be applied to any linear regression technique.

The best linear coefficients $\boldsymbol{w}$ enabled not only the prediction of output data (jumping height) but also the dissociation of input data into a task-relevant component $\boldsymbol{X}_{\mathrm{rel}}$ and task-irrelevant component $\boldsymbol{X}_{\mathrm{irr}}$. By minimizing the cost function

$$E_{\mathrm{rel}} = \frac{1}{2}(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{X}_{\mathrm{rel}}\boldsymbol{w})^T(\boldsymbol{X}\boldsymbol{w} - \boldsymbol{X}_{\mathrm{rel}}\boldsymbol{w}) \tag{1}$$

under the constraint $\boldsymbol{X} \neq \boldsymbol{X}_{\mathrm{rel}}$ (avoidance of a self-evident answer), $\boldsymbol{X}_{\mathrm{rel}}$ can be written as

$$\boldsymbol{X}_{\mathrm{rel}} = \boldsymbol{X}\boldsymbol{w}\boldsymbol{w}^T(\boldsymbol{w}\boldsymbol{w}^T)^{\dagger} = \boldsymbol{X}\frac{\boldsymbol{w}\boldsymbol{w}^T}{|\boldsymbol{w}|^2}, \tag{2}$$

where $(\boldsymbol{w}\boldsymbol{w}^T)^{\dagger}$ is a pseudo-inverse of $\boldsymbol{w}\boldsymbol{w}^T$ and $|\boldsymbol{w}| = \sqrt{\boldsymbol{w}^T\boldsymbol{w}}$. The equality $\boldsymbol{w}^T(\boldsymbol{w}\boldsymbol{w}^T)^{\dagger} = \frac{\boldsymbol{w}^T}{|\boldsymbol{w}|^2}$ holds when $\boldsymbol{w} \in \mathrm{R}^{D\times 1}$. Under the decomposition $\boldsymbol{X} = \boldsymbol{X}_{\mathrm{rel}} + \boldsymbol{X}_{\mathrm{irr}}$, $\boldsymbol{X}_{\mathrm{irr}}$ can be written as

$$\boldsymbol{X}_{\mathrm{irr}} = \boldsymbol{X} - \boldsymbol{X}_{\mathrm{rel}} = \boldsymbol{X}\left(\boldsymbol{I} - \frac{\boldsymbol{w}\boldsymbol{w}^T}{|\boldsymbol{w}|^2}\right), \tag{3}$$

where $\boldsymbol{I} \in \mathrm{R}^{D\times D}$ is an identity matrix.

Under the condition $\boldsymbol{X} = \boldsymbol{X}_{\mathrm{rel}} + \boldsymbol{X}_{\mathrm{irr}}$, the variance of the ith component of $\boldsymbol{X}$, X_i, can be calculated as

$$\mathrm{Var}(X_i) = \frac{1}{T}\sum_{t=1}^{T} X_{i,t}^2 = \frac{1}{T}\sum_{t=1}^{T}(X_{i,t}^{\mathrm{rel}} + X_{i,t}^{\mathrm{irr}})^2 = \mathrm{Var}(X_i^{\mathrm{rel}}) + \mathrm{Var}(X_i^{\mathrm{irr}}) + 2\mathrm{Cov}(X_i^{\mathrm{rel}}, X_i^{\mathrm{irr}}), \tag{4}$$

where $X_{i,t}$ is X_i at the tth trial, $X_{i,t}^{\mathrm{rel}}$ is the ith component of $\boldsymbol{X}^{\mathrm{rel}}$ at the tth trial, $X_{i,t}^{\mathrm{irr}}$ is the ith component of $\boldsymbol{X}^{\mathrm{irr}}$ at the tth trial, and $\mathrm{Cov}(X_i^{\mathrm{rel}}, X_i^{\mathrm{irr}})$ is the covariance between X_i^{rel} and X_i^{irr}. Notably, in the current experimental setting, $\mathrm{Cov}(X_i^{\mathrm{rel}}, X_i^{\mathrm{irr}})$ in the analyzed trials was close to 0. We thus considered only $\mathrm{Var}(X_i^{\mathrm{rel}})$ and $\mathrm{Var}(X_i^{\mathrm{irr}})$.

Ridge regression The ridge regression enabled us to determine the best one-dimensional linear space $\boldsymbol{w} \in \mathrm{R}^{D\times 1}$ in the input data $\boldsymbol{X} \in \mathrm{R}^{T\times D}$ to predict the output data $\boldsymbol{y} \in \mathrm{R}^{T\times 1}$ by minimizing the cost function:

$$E = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) + \frac{\lambda}{2}\boldsymbol{w}^T\boldsymbol{w}. \tag{5}$$

The first term on the right-hand side indicates the fitting error, the second term indicates the regularization of $\boldsymbol{w}$, and λ is a regularization parameter. The current study determined λ to minimize the prediction error based on a 10-fold cross validation, which enabled us to avoid overfitting [29]. Overfitting, which can appear without any regularization, leads to the selection of a model that is more complicated than the true one. Minimization of the cost function concerning $\boldsymbol{w}$ leads to the optimal value for $\boldsymbol{w}$:

$$\boldsymbol{w}^* = \left(\boldsymbol{X}\boldsymbol{X}^T + \lambda\boldsymbol{I}\right)^{-1}\boldsymbol{X}^T\boldsymbol{y}, \tag{6}$$

where $\boldsymbol{I}$ was an identity matrix.

The ridge regression enabled the estimation of an appropriate $\boldsymbol{w}$ based on the normalized $\boldsymbol{y}$ and $\boldsymbol{X}$, i.e., the mean and standard deviation of $\boldsymbol{y}$ and $\boldsymbol{X}$ should be normalized to be 0 and 1, respectively; $\frac{1}{T}\sum_{t=1}^{T} y_t = 0$, $\frac{1}{T}\sum_{t=1}^{T} y_t^2 = 1$, $\frac{1}{T}\sum_{t=1}^{T} X_{i,t} = 0$, and $\frac{1}{T}\sum_{t=1}^{T} X_{i,t}^2 = 1$ $(d = 1, ..., D)$. All the results in the current study depended on the normalized data. Without normalization, w_i is estimated to be large when $X_{i,t}$ shows small fluctuations and vice versa, although regularization with parameter λ was imposed

equally to all the w_i; therefore, normalization, especially in $\boldsymbol{X}$, is indispensable for estimating appropriate $\boldsymbol{w}$. Notably, the normalization did not affect interpretation at all because it was possible to restore the original unnormalized data by adding the original mean $m_i = \frac{1}{T}\sum_{t=1} X_{i,t}^{\text{original}}$ and multiplying the original standard deviation $s_i = \sqrt{\frac{1}{T}\sum_{t=1}(X_{i,t}^{\text{original}})^2 - m_d^2}$. To satisfy $\sum_{i=1}^{D} w_i X_{i,t} = \sum_{i=1}^{D} \tilde{w}_i X_{i,t}^{\text{original}}$, $\tilde{\boldsymbol{w}}$, where $\boldsymbol{w}$ corresponds to unnormalized data, should be divided by s_i ($\tilde{w}_i = \frac{w_i}{s_i}$) and $\sum_{i=1}^{D} \tilde{w}_i m_i$ should be subtracted. In total, the normalization is indispensable for estimating an appropriate $\boldsymbol{w}$; however, it did not affect the results at all.

The ridge regression showed high prediction power under the existence of measurement noise in $\boldsymbol{X}$. Under the existence of measurement Gaussian noise $\boldsymbol{\xi}$ with a mean of 0, the standard deviation is σ, covariance is 0, and the cost function averaged across all the possible noise can be written as

$$\langle E \rangle = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}) + \frac{\sigma^2}{2}\boldsymbol{w}^T\boldsymbol{w}. \tag{7}$$

The equivalence between equations (5) and (7) indicates that the ridge regression enabled the selection of the best $\boldsymbol{w}$ to predict $\boldsymbol{y}$ under the existence of measurement noise while avoiding overfitting.

Parabolic representation of jumping height, three candidates of input data, the UCM and GEM The vertical position of the marker attached to the subject' s back determined the jumping height in the current study. We expected that the jumping height could be predicted well based on the position p and velocity v of the marker at the release timing as follows

$$h_t = p_t + \frac{v_t^2}{2g}, \tag{8}$$

where $g \simeq 9.8(m/s^2)$. In the current setting, the markers attached to the subject' s back and hip joint showed similar velocities (i.e., the joint angle between the hip joint and back equaled close to 90° at the release timing); it was thus possible to predict the jumping height based on the position and velocity of the marker attached to the hip joint and the release timing, although there was a constant bias between the position of the markers attached to the back and hip. In other words, the position and velocity of the hip joint were factors relevant to task performance in the current study. In the joint angle representation, p and v were written as follows:

$$p = \sum_{i=1}^{3} l_i \sin q_i \tag{9}$$

and

$$v = \sum_{i=1}^{3} l_i \dot{q}_i \cos q_i. \tag{10}$$

In the UCM (blue crosses in Fig. 4A), we calculated the task-relevant and task-irrelevant variabilities based on equations (9) and (10).

Using the equations (9) and (10), the predicted jumping height h_t can be written as

$$h_t = \sum_{i=1}^{3} l_i \sin q_{i,t} + \frac{1}{2g}\sum_{i=1}^{3}\sum_{j=1}^{3} l_i l_j \dot{q}_{i,t}\dot{q}_{j,t} \cos q_{i,t} \cos q_{j,t}. \tag{11}$$

The first candidate input data for the ridge regression were the joint angles and angular velocities (blue line in Fig. 3A). The second candidate data were the functions in the forward kinematics of the position and velocity of the hip joint (equations (9) and (10), red line in Fig. 3A). The third candidate data were the functions that appeared in equation (11) (orange line in Fig. 3A). In GEM (blue crosses in Fig. 4A), we calculated the task-relevant and task-irrelevant variabilities based on equation (11).

Relation between ridge regression and principal component analysis It was possible to analytically determine the relation between the ridge regression and principal component analysis (PCA) by decomposing $\boldsymbol{X}$ using singular value decomposition (SVD), $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T$, where $\boldsymbol{U} \in \mathbf{R}^{T\times T}$ is an orthogonal matrix, $\boldsymbol{D} \in \mathbf{R}^{T\times D}$ includes the square root of the ith eigenvalue of $\boldsymbol{X}^T\boldsymbol{X}$ at (i,i) element and $D_{i,j} = 0$ when $i \neq j$, and $\boldsymbol{V} \in \mathbf{R}^{D\times D}$ is an orthogonal matrix. Using the SVD and equation (6), the predicted output h_t can be written as

$$h_t = \boldsymbol{X}_t\boldsymbol{w}^* = \boldsymbol{U}\boldsymbol{D}(\boldsymbol{D}^T\boldsymbol{D} + \lambda\boldsymbol{I})^{-1}\boldsymbol{D}^T\boldsymbol{U}^T\boldsymbol{y} = \sum_{i=1}^{\min(T,D)} \frac{\lambda_i^2}{\lambda_i^2 + \lambda}\mathrm{Corr}(\boldsymbol{X}\boldsymbol{v}_i, \boldsymbol{y})u_{i,t}, \tag{12}$$

where $\min(T,D)$ determines the rank of $\boldsymbol{X}$, λ_i^2 is an eigenvalue of $\boldsymbol{X}^T\boldsymbol{X}$, $\mathrm{Corr}(\cdot,\cdot)$ indicates the correlation between two vectors, $\boldsymbol{v}_i$ is the eigenvector of $\boldsymbol{X}^T\boldsymbol{X}$ corresponding to λ_i^2, and $u_{i,t}$ is the (i,t) component of $\boldsymbol{U}$. On the other hand, PCA enables the decomposition of $\boldsymbol{X}_t$ as

$$\boldsymbol{X}_t = \sum_{i=1}^{\min(T,D)} \lambda_i u_{i,t}\boldsymbol{v}_i. \tag{13}$$

This equation indicates that the motion data can be decomposed into eigenvectors (principal components) with weight $\lambda_i u_{i,t}$. By comparing equations (12) and (13), the ridge regression enables the prediction of output data by weighting based on the ith eigenvector with weight $\frac{\lambda_i^2}{\lambda_i^2+\lambda}u_{i,t}$ (notably, $\frac{\lambda_i^2}{\lambda_i^2+\lambda}$ was a monotonic function concerning λ_i). An important difference between PCA and the ridge regression is whether the task relevance of the ith eigenvector, $\mathrm{Corr}(\boldsymbol{X}\boldsymbol{v}_i, \boldsymbol{y})$, should be considered. Although PCA relies only on the eigenvalue, the ridge regression considers both (nonlinearly transformed) eigenvalue and task relevance. The ridge regression could thus be considered an extended version of PCA to determine how each principal component is relevant to the task.

In PCA, we found the relation between the explained variance and prediction power to be as follows: at a s% explained variance, we determine the number of principal components based on $n_s = \min_n \sum_{i=1}^{n} \lambda_i^2 > s$ (i.e., the minimum number of principal components that exceed s% explained variance). After determining n_s, motion data can be reconstructed as $\tilde{\boldsymbol{X}}_t = \sum_{i=1}^{n_s} \lambda_i u_{i,t}\boldsymbol{v}_i$. We then multiply $\sum_{i=1}^{n_s} \boldsymbol{v}_i^T$ by the $\tilde{\boldsymbol{X}}_t$ from the right-hand side, resulting in $\tilde{y}_t = \tilde{\boldsymbol{X}}_t \sum_{i=1}^{n_s} \boldsymbol{v}_i^T = \sum_{i=1}^{n_s} \lambda_i u_{i,t}$. Finally, we calculate the correlation between observed jumping height h_t and $\tilde{y}_t$ in Figs. 4C-4E.

Acknowledgement

This work was supported by a Grant-in-Aid for Young Scientists (18K17894). We thank D. Nozaki and S. Hagio for their helpful comments.

Additional Information

Competing financial interests: The authors declare no competing financial interests.

Author contributions

D.F. and K.T. designed and performed the experiments. D.F. and K.T. performed the analyses and wrote the manuscript.

References

[1] Bernstein, N. A. The co-ordination and regulation of movements. Pergamon, London (1967).

[2] Faisal, A. A., Selen, L. P. J. & Wolpert, D. M. Noise in the nervous system. Nature Neuroscience 9, 292-303 (2008).

[3] Churchland, M. M., Afshar, A. & Shenoy, K. V. A Central Source of Movement Variability. Neuron 52, 1085-1096 (2006).

[4] Jones, K. E., Hamilton, A. F. & Wolpert, D. M. Sources of signal-dependent noise during isometric force production. Journal of Neurophysiology 88, 1533-1544 (2002).

[5] Dhawale, A. K., Smith, M. A. & Ölveczky, B. P. The Role of Variability in Motor Learning. Annual Review of Neuroscience 40, 479-498 (2017).

[6] Bizzi, E., Mussa-Ivaldi, F. A. & Giszter, S. Computations underlying the execution of movement: a biological perspective. Science 253, 287-291 (1991).

[7] Gribble, P. L., Mullin, L. I., Cothros, N. & Mattar, A. Role of cocontraction in arm movement accuracy. Journal of Neurophysiology 89, 2396-2405 (2003).

[8] Latash, M. L., Scholz, J. P. & Sch ö ner, G. Motor control strategies revealed in the structure of motor variability. Exercise and sport sciences reviews 30, 26-31 (2002).

[9] Rokni, U., Richardson, A. G., Bizzi, E. & Seung, H. S. Motor Learning with Unstable Neural Representations. Neuron 54, 653-666 (2007).

[10] Takiyama, K. & Okada, M. Maximization of Learning Speed in the Motor Cortex Due to Neuronal Redundancy. PLoS Computational Biology 8, e1002348 (2012).

[11] Scholz, J. P. & Sch ö ner, G. The uncontrolled manifold concept: identifying control variables for a functional task. Experimental Brain Research 126, 289-306 (1999).

[12] Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. Nature Neuroscience 5, 1226-1235 (2002).

[13] Müller, H. & Sternad, D. A randomization method for the calculation of covariation in multiple nonlinear relations: illustrated with the example of goal-directed movements. Biological Cybernetics 89, 22-33 (2003).

[14] Cusumano, J. P. & Cesari, P. Body-goal variability mapping in an aiming task. Biological Cybernetics 94, 367-379 (2006).

[15] Kaufman, M. T., Churchland, M. M., Ryu, S. I. & Shenoy, K. V. Cortical activity in the null space: permitting preparation without movement. Nature Neuroscience 1-12 (2014).

[16] Krishnamoorthy, V., Latash, M. L., Scholz, J. P. & Zatsiorsky, V. M. Muscle synergies during shifts of the center of pressure by standing persons. Experimental Brain Research 152, 281-292 (2003).

[17] Valero-Cuevas, F. J., Venkadesan, M. & Todorov, E. Structured Variability of Muscle Activations Supports the Minimal Intervention Principle of Motor Control. Journal of Neurophysiology 102, 59-68 (2009).

[18] Yang, J.-F., Scholz, J. P. & Latash, M. L. The role of kinematic redundancy in adaptation of reaching. Experimental Brain Research 176, 54-69 (2006).

[19] d'Avella, A., Saltiel, P. & Bizzi, E. Combinations of muscle synergies in the construction of a natural motor behavior. Nature Neuroscience 6, 300-308 (2003).

[20] Ivanenko, Y. P., Cappellini, G., Dominici, N., Poppele, R. E. & Lacquaniti, F. Modular control of limb movements during human locomotion. The Journal of Neuroscience 27, 11149-11161 (2007).

[21] Hagio, S. & Kouzaki, M. The flexible recruitment of muscle synergies depends on the required force-generating capability. Journal of Neurophysiology 112, 316-327 (2014).

[22] Furuya, S., Tominaga, K., Miyazaki, F. & Altenmüller, E. Losing dexterity: patterns of impaired coordination of finger movements in musician' s dystonia. Scientific Reports 5, 13360 (2015).

[23] Lacquaniti, F. & Maioli, C. Independent control of limb position and contact forces in cat posture. Journal of Neurophysiology 72, 1476-1495 (1994).

[24] Dominici, N. et al. Locomotor Primitives in Newborn Babies and Their Development. Science 334,

997-999 (2011).
[25] Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788-791 (1999).
[26] Latash, M. L., Scholz, J. P. & Schöner, G. Motor control strategies revealed in the structure of motor variability. Exercise and sport sciences reviews 30, 26-31 (2002).
[27] Santello, M. & Soechting, J. F. Gradual molding of the hand to object contours. Journal of Neurophysiology 79, 1307-1320 (1998).
[28] Delis, I., Berret, B., Pozzo, T. & Panzeri, S. Quantitative evaluation of muscle synergy models: a single-trial task decoding approach. Frontiers in Computational Neuroscience 7 (2013).
[29] Bishop, C. M. Pattern Recognition and Machine Learning (Springer Verlag, 2006).
[30] Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics (1970).
[31] Furuki, D. & Takiyama, K. Detecting the relevance to performance of whole-body movements. Scientific Reports 7, 15659 (2017).
[32] Krakauer, J. W., Pine, Z. M., Ghilardi, M. F. & Ghez, C. Learning of visuomotor transformations for vectorial planning of reaching trajectories. The Journal of Neuroscience 20, 8916-8924 (2000).
[33] Cotti, J. et al. Adaptation of reactive and voluntary saccades: different patterns of adaptation revealed in the antisaccade task. The Journal of Physiology 587, 127-138 (2009).
[34] Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., lveczky, B. P. O. & Smith, M. A. Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. Nature Neuroscience 1-13 (2014).
[35] Kagerer, F. A., Contreras-Vidal, J. L. & Stelmach, G. E. Adaptation to gradual as compared with sudden visuo-motor distortions. Experimental Brain Research 115, 557-561 (1997).
[36] Sternad, D., Park, S.-W., Mueller, H. & Hogan, N. Coordinate Dependence of Variability Analysis. PLoS Computational Biology 6 (2010).
[37] Shinya, M. et al. Pitching form determines probabilistic structure of errors in pitch location. Journal of sports sciences 35, 2142-2147 (2017).
[38] Thoroughman, K. A. & Shadmehr, R. Learning of action through adaptive combination of motor primitives . Nature 407, 742-747 (2000).
[39] Scheidt, R. A., Dingwell, J. B. & Mussa-Ivaldi, F. A. Learning to move amid uncertainty. Journal of Neurophysiology 86, 971-985 (2001).
[40] Donchin, O., Francis, J. T. & Shadmehr, R. Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. The Journal of Neuroscience 23, 9032-9045 (2003).
[41] Smith, M. A., Ghazizadeh, A. & Shadmehr, R. Interacting adaptive processes with different timescales underlie short-term motor learning. PLoS Biology 4, e179 (2006).
[42] Takiyama, K., Hirashima, M. & Nozaki, D. Prospective errors determine motor learning. Nature Communications 6, 5925 (2015).
[43] Takiyama, K. Context-dependent memory decay is evidence of effort minimization in motor learning: a computational study. Frontiers in Computational Neuroscience 9, 1-10 (2015).
[44] Takiyama, K. & Sakai, Y. Balanced motor primitive can explain generalization of motor learning effects between unimanual and bimanual movements. Scientific Reports 1-10 (2016).
[45] Takiyama, K. & Sakai, Y. A balanced motor primitive framework can simultaneously explain motor learning in unimanual and bimanual movements. Neural Networks 86, 80-89 (2017).
[46] Takiyama, K., Katahira, K. & Okada, M. Exact inference in discontinuous firing rate estimation using belief propagation. Journal of the Physical Society of Japan 78, 4003 (2009).
[47] Takiyama, K. & Okada, M. Detection of hidden structures in nonstationary spike trains. Neural Computation 23, 1205-1233 (2011).
[48] Naruse, Y., Takiyama, K., Okada, M. & Umehara, H. Statistical method for detecting phase shifts

in alpha rhythm from human electroencephalogram data. Physical review. E, Statistical, nonlinear, and soft matter physics 87, 042708 (2013).
[49] Tibshirani, R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B (1996).
[50] Rasmussen, C. E. & Williams, C. K. I. Gaussian Processes for Machine Learning (MIT Press, 2006).

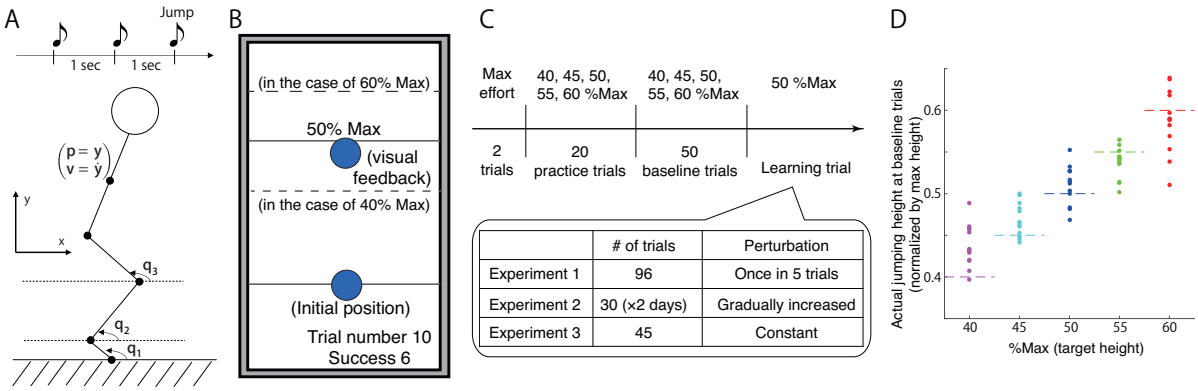


Figure 1. Summary of our experimental settings. **A**: Participants were instructed to perform a vertical jump at the timing of the third beep. Three beeps sounded at one-second intervals. We measured and analyzed joint angles at toe, ankle, and knee in the sagittal plane. The jumping height was measured based on the position of the marker attached to the back in the y-axis. **B**: Task instruction and feedback information in each trial. A computer monitor was located in front of the participants (1.5 meters ahead, 1.7 meters above the floor). One second before the first beep, target height (indicated by black bar and texts [e.g., 50% max]), baseline height (indicated by black bar), and initial position (indicated by blue cursor located on the baseline height) were displayed. When the target height was 60%, the black bar and text were displayed at the position of the higher black dotted bar. When the target height was 40%, the black bar and text were displayed at the position of the lower black dotted bar. These black dotted bars were used only for the explanation and were not visible throughout the experiments. In the practice trials, the blue cursor was displayed during trials to continuously indicate the position of the marker attached to the back in the y-axis. These trials enabled the participants to become accustomed to the experimental setting. In baseline and learning trials, the blue cursor was displayed at the beginning and end of each trial. At the beginning of each trial, the blue cursor was displayed at the baseline height. At the end of each trial, the cursor was displayed depending on the actual jumping height. When the jumping height was close to the target height, the participants heard a coin-getting sound. During the experiments, the subjects were provided with the current trial number and the number of successful trials where they could hear change rattling. **C**: The sequence of the experiments. Participants performed a vertical jump with maximum effort for two trials. These jumping heights were used to determine the target height. Participants experienced 20 practice trials, 50 baseline trials, and a number of learning trials specific to each experiment. **D**: Averaged jumping height of each participant in the baseline trials in experiment 1. The jumping height depended on the target height (1-way ANOVA, $p = 2.2597 \times 10^{-21}$), indicating that the participants could perform the goal-directed movement.

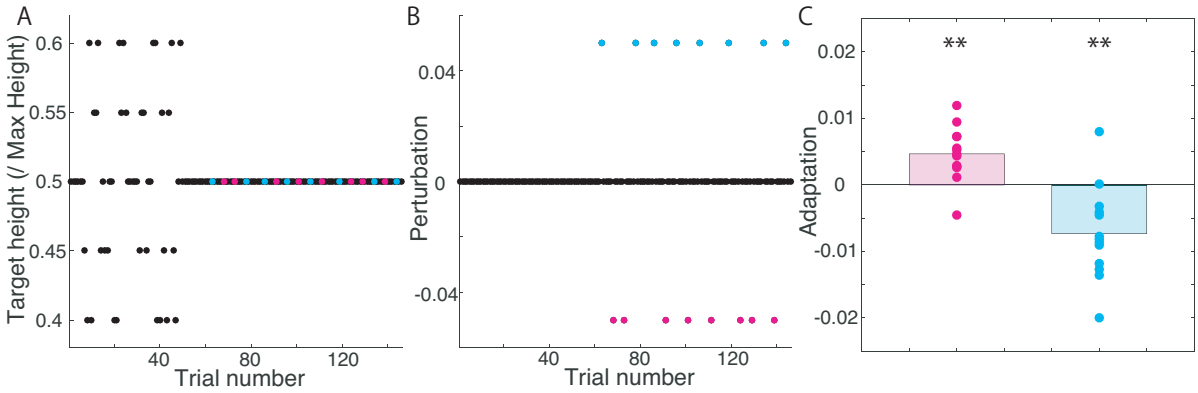


Figure 2. Diagram and results of experiment 1. **A**: Target height in baseline and learning trials. Cyan and magenta circles indicated the trials with perturbations. **B**: Perturbation sequence. The cyan circle indicates the trials with p=0.05, and the magenta circle indicates those with $p = -0.05$. The perturbations were pseudorandomly imposed once in five trials. **C**: Adaptation effect. The vertical line indicates the modification of the jumping height after the perturbation was imposed. Magenta dots indicate the averaged difference in each subject corresponding to the perturbation $p = -0.05$, and cyan dots indicate the averaged difference in each subject corresponding to the perturbation $p = 0.05$.

Figure legends

Figure 1. Summary of our experimental settings. **A**: Participants were instructed to perform a vertical jump at the timing of the third beep. Three beeps sounded at one-second intervals. We measured and analyzed joint angles at toe, ankle, and knee in the sagittal plane. The jumping height was measured based on the position of the marker attached to the back in the y-axis. **B**: Task instruction and feedback information in each trial. A computer monitor was located in front of the participants (1.5 meters ahead, 1.7 meters above the floor). One second before the first beep, target height (indicated by black bar and texts [e.g., 50% max]), baseline height (indicated by black bar), and initial position (indicated by blue cursor located on the baseline height) were displayed. When the target height was 60%, the black bar and text were displayed at the position of the higher black dotted bar. When the target height was 40%, the black bar and text were displayed at the position of the lower black dotted bar. These black dotted bars were used only for the explanation and were not visible throughout the experiments. In the practice trials, the blue cursor was displayed during trials to continuously indicate the position of the marker attached to the back in the y-axis. These trials enabled the participants to become accustomed to the experimental setting. In baseline and learning trials, the blue cursor was displayed at the beginning and end of each trial. At the beginning of each trial, the blue cursor was displayed at the baseline height. At the end of each trial, the cursor was displayed depending on the actual jumping height. When the jumping height was close to the target height, the participants heard a coin-getting sound. During the experiments, the subjects were provided with the current trial number and the number of successful trials where they could hear change rattling. **C**: The sequence of the experiments. Participants performed a vertical jump with maximum effort for two trials. These jumping heights were used to determine the target height. Participants experienced 20 practice trials, 50 baseline trials, and a number of learning trials specific to each experiment. **D**: Averaged jumping height of each participant in the baseline trials in experiment 1. The jumping height depended on the target height (1-way ANOVA, $p = 2.2597 \times 10^{-21}$), indicating that the participants could perform the goal-directed movement.

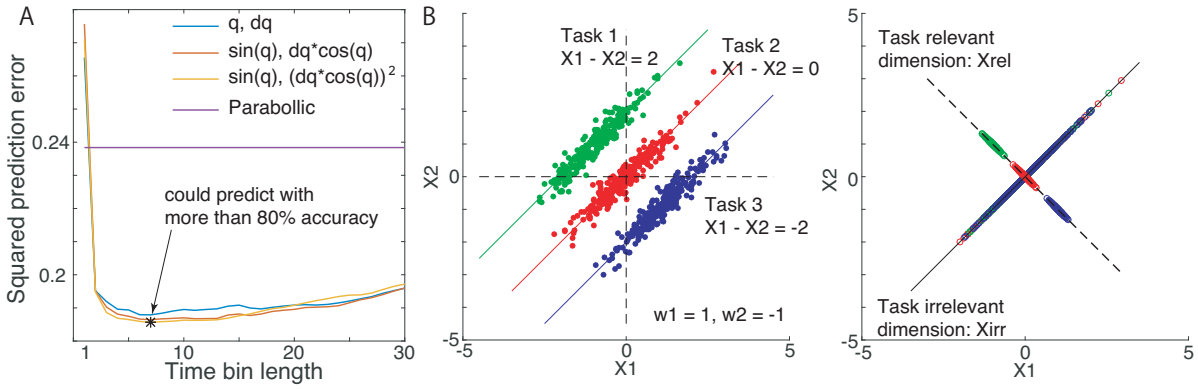


Figure 3. Validation of the ridge regression and the concept of our method. **A**: Predictive power of the ridge regression using three kinds of input data. Horizontal and vertical axes indicate the time bin length used for the ridge regression and squared prediction error, respectively. If the ridge regression could not make a prediction, the prediction error equaled 1. If the ridge regression could predict the output data perfectly, the prediction error equaled 0. These results indicate that the ridge regression enables the prediction of output data with a greater than 80% accuracy. **B**: An example of decomposing input data into task-relevant and task-irrelevant components. In this case, we assumed that the task 1 required $X_1 - X_2$ to be 2 (green line), task 2 required $X_1 - X_2$ to be 0 (green line), and task 3 required $X_1 - X_2$ to be -2 (green line). Green, red, and blue dots indicate the typical input data for tasks 1, 2, and 3, respectively. In the ridge regression, these tasks can be achieved with $w_1 = 1$ and $w_2 = -1$, i.e., $y = w_1 X_1 + w_2 X_2 = X_1 - X_2$ should be determined differently in each task. In the right panel, these input data were decomposed into a task-relevant (black dotted line) component $X_{\mathrm{rel}} = X w w^T / |w|^2$ and a task-irrelevant component $X_{\mathrm{irr}} = X - X_{\mathrm{rel}}$ (solid black line). X_{rel} was separated depending on the task, and X_{irr} was not separated, which indicates that the decomposition enables the discussion of the task-relevant and task-irrelevant components.

Figure 2. Diagram and results of experiment 1. **A**: Target height in baseline and learning trials. Cyan and magenta circles indicated the trials with perturbations. **B**: Perturbation sequence. The cyan circle indicates the trials with p=0.05, and the magenta circle indicates those with $p = -0.05$. The perturbations were pseudorandomly imposed once in five trials. **C**: Adaptation effect. The vertical line indicates the modification of the jumping height after the perturbation was imposed. Magenta dots indicate the averaged difference in each subject corresponding to the perturbation $p = -0.05$, and cyan dots indicate the averaged difference in each subject corresponding to the perturbation $p = 0.05$.

Figure 3. Validation of the ridge regression and the concept of our method. **A**: Predictive power of the ridge regression using three kinds of input data. Horizontal and vertical axes indicate the time bin length used for the ridge regression and squared prediction error, respectively. If the ridge regression could not make a prediction, the prediction error equaled 1. If the ridge regression could predict the output data perfectly, the prediction error equaled 0. These results indicate that the ridge regression enables the prediction of output data with a greater than 80% accuracy. **B**: An example of decomposing input data into task-relevant and task-irrelevant components. In this case, we assumed that the task 1 required $X_1 - X_2$ to be 2 (green line), task 2 required $X_1 - X_2$ to be 0 (green line), and task 3 required $X_1 - X_2$ to be -2 (green line). Green, red, and blue dots indicate the typical input data for tasks 1, 2, and 3, respectively. In the ridge regression, these tasks can be achieved with $w_1 = 1$ and $w_2 = -1$, i.e., $y = w_1 X_1 + w_2 X_2 = X_1 - X_2$ should be determined differently in each task. In the right panel, these input data were decomposed into a task-relevant (black dotted line) component $X_{\mathrm{rel}} = X w w^T / |w|^2$ and

a task-irrelevant component $X_{\mathrm{irr}} = X - X_{\mathrm{rel}}$ (solid black line). X_{rel} was separated depending on the task, and X_{irr} was not separated, which indicates that the decomposition enables the discussion of the task-relevant and task-irrelevant components.

Figure 4. Validation of our method and comparison to previous methods. **A**: Evaluation of task-relevant and task-irrelevant variabilities. Red dots indicate those variabilities evaluated by our method in each subject (N = 13). Blue and green crosses indicate the variabilities evaluated by the UCM and GEM, respectively. Our method uses a different normalization method from those of the UCM and GEM. **B**: Correlation between predicted and actual jumping height. The red line and the shaded area indicate the mean and standard error of the mean (s.e.m.) of the correlation in ridge regression (N = 13), respectively. The blue line and shaded area indicate the mean and s.e.m. of the correlation in PCA (N = 13), respectively. Horizontal and vertical axes indicate the explained variance or the corresponding number of principal components and the correlation, respectively. **C**: Explained variance and the correlation between the predicted and actual jumping height of each principal component. The red line and the shaded area indicate the mean and s.e.m. of the correlation (N = 13), respectively. The blue line and the shaded area indicate mean and s.e.m. of the variance explained (N = 13), respectively. **D**: Variance explained and the correlation between the predicted and actual jumping height of each principal component in a typical subject. The red and blue lines indicate the correlation and the variance explained, respectively. **E**: The PC number that is most relevant to predicting jumping height. Horizontal and vertical axes indicate the subject number and the most relevant PC number, respectively.

Figure 5. Diagram and results of experiments 2 and 3. **A,D**: Target height in baseline and learning trials. Blue and red circles indicate the trials with perturbations. **B,E**: Perturbation sequence. Subjects joined experiment 2 for two days and experienced two different perturbations (either $p > 0$ or $p < 0$). The order of perturbation was counterbalanced across subjects. In experiment 3, subjects experienced both positive and negative perturbations within one day. Although panel E shows the case when a negative perturbation followed a positive one, the order of the perturbation was counterbalanced across subjects. **C,F**: Learning curves. The thin solid lines indicate the learning curves of each subject. The bold solid line indicates the learning curve averaged across all subjects.

Figure 6. Application of our method to the results of experiments 2 and 3. **A**: Task-relevant variabilities of each subject (N = 13) before and after adaptation to perturbation in experiments 2 and 3. The blue and red dots indicate the variability of each subject in experiments 2 and 3, respectively. The blue and red lines indicate the modulation of the variability due to adaptation. The blue and red bars indicate the averaged variability across all subjects. There was a significant difference between the variabilities before the adaptation and those after the adaptation in experiment 3 (Wilcoxon signed rank test, p = 0.0105). **B**: Task-irrelevant variabilities of each subject (N = 13) before and after adaptation to perturbation in experiments 2 and 3. There was a significant difference between the variabilities before the adaptation and those after the adaptation in experiment 2 (Wilcoxon signed rank test, p = 0.0215). **C,D**: Interpretation of our results based on a simple example. We assume that the task before adaptation required $X_1 - X_2$ to be 2 and that the task after adaptation required $X_1 - X_2$ to be -2. Panel C indicates an interpretation of our results in experiment 2. In the experiment, task-irrelevant variabilities increased after adaptation, and task-relevant variabilities remained unchanged. Panel D indicates an interpretation of our results from experiment 3. In the experiment, task-relevant variabilities increased after adaptation, and task-irrelevant variabilities remained unchanged.

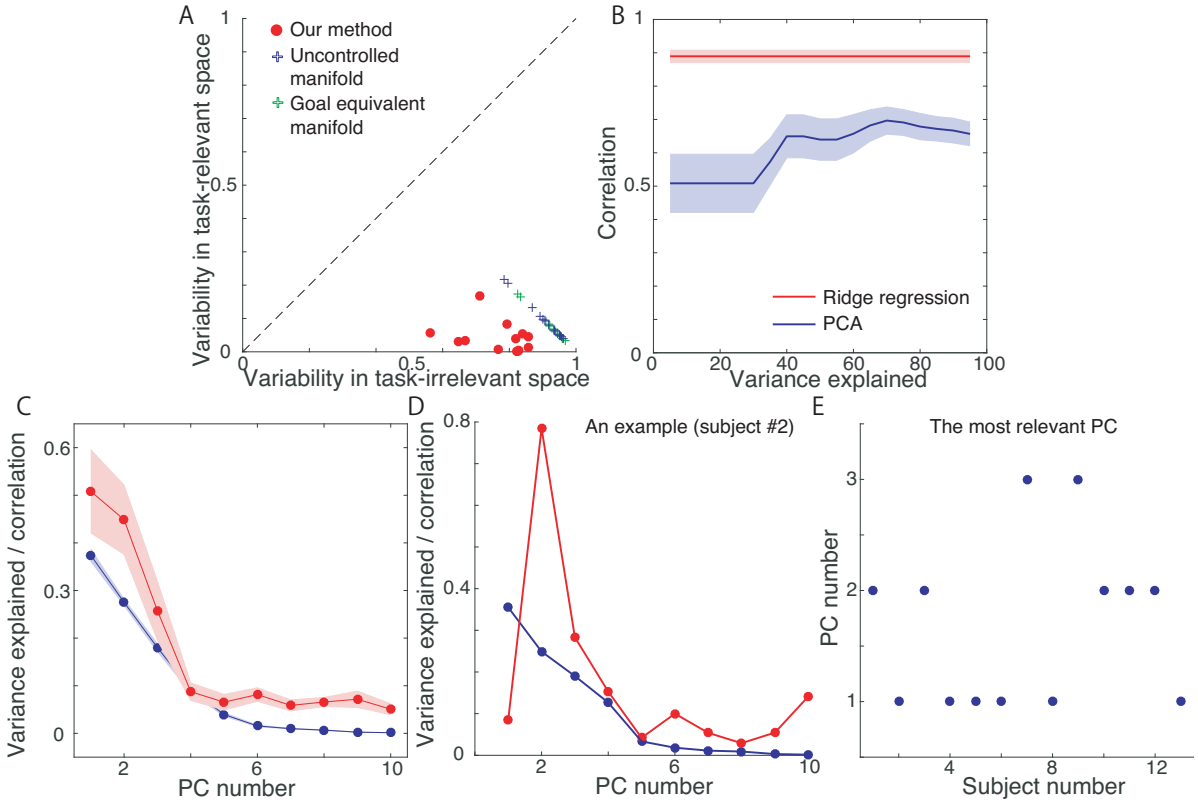


Figure 4. Validation of our method and comparison to previous methods. **A**: Evaluation of task-relevant and task-irrelevant variabilities. Red dots indicate those variabilities evaluated by our method in each subject (N = 13). Blue and green crosses indicate the variabilities evaluated by the UCM and GEM, respectively. Our method uses a different normalization method from those of the UCM and GEM. **B**: Correlation between predicted and actual jumping height. The red line and the shaded area indicate the mean and standard error of the mean (s.e.m.) of the correlation in ridge regression (N = 13), respectively. The blue line and shaded area indicate the mean and s.e.m. of the correlation in PCA (N = 13), respectively. Horizontal and vertical axes indicate the explained variance or the corresponding number of principal components and the correlation, respectively. **C**: Explained variance and the correlation between the predicted and actual jumping height of each principal component. The red line and the shaded area indicate the mean and s.e.m. of the correlation (N = 13), respectively. The blue line and the shaded area indicate mean and s.e.m. of the variance explained (N = 13), respectively. **D**: Variance explained and the correlation between the predicted and actual jumping height of each principal component in a typical subject. The red and blue lines indicate the correlation and the variance explained, respectively. **E**: The PC number that is most relevant to predicting jumping height. Horizontal and vertical axes indicate the subject number and the most relevant PC number, respectively.

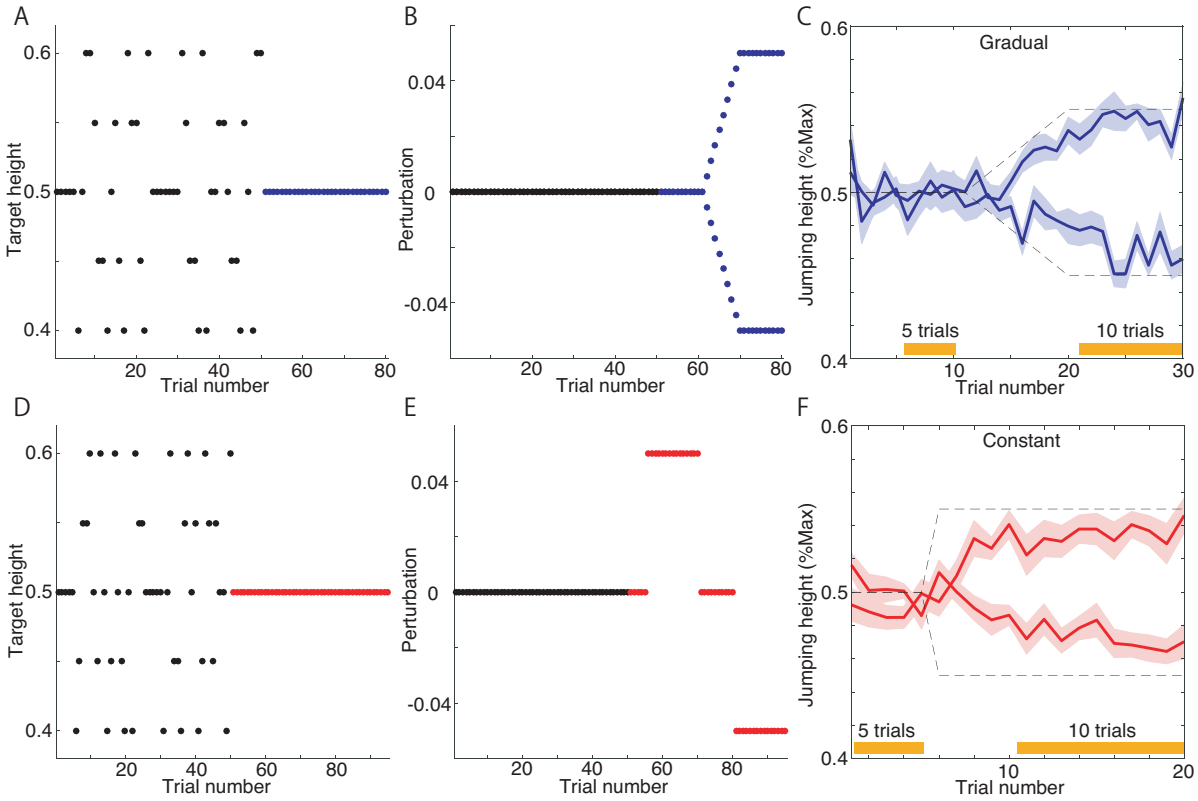


Figure 5. Diagram and results of experiments 2 and 3. **A,D**: Target height in baseline and learning trials. Blue and red circles indicate the trials with perturbations. **B,E**: Perturbation sequence. Subjects joined experiment 2 for two days and experienced two different perturbations (either $p > 0$ or $p < 0$). The order of perturbation was counterbalanced across subjects. In experiment 3, subjects experienced both positive and negative perturbations within one day. Although panel E shows the case when a negative perturbation followed a positive one, the order of the perturbation was counterbalanced across subjects. **C,F**: Learning curves. The thin solid lines indicate the learning curves of each subject. The bold solid line indicates the learning curve averaged across all subjects.

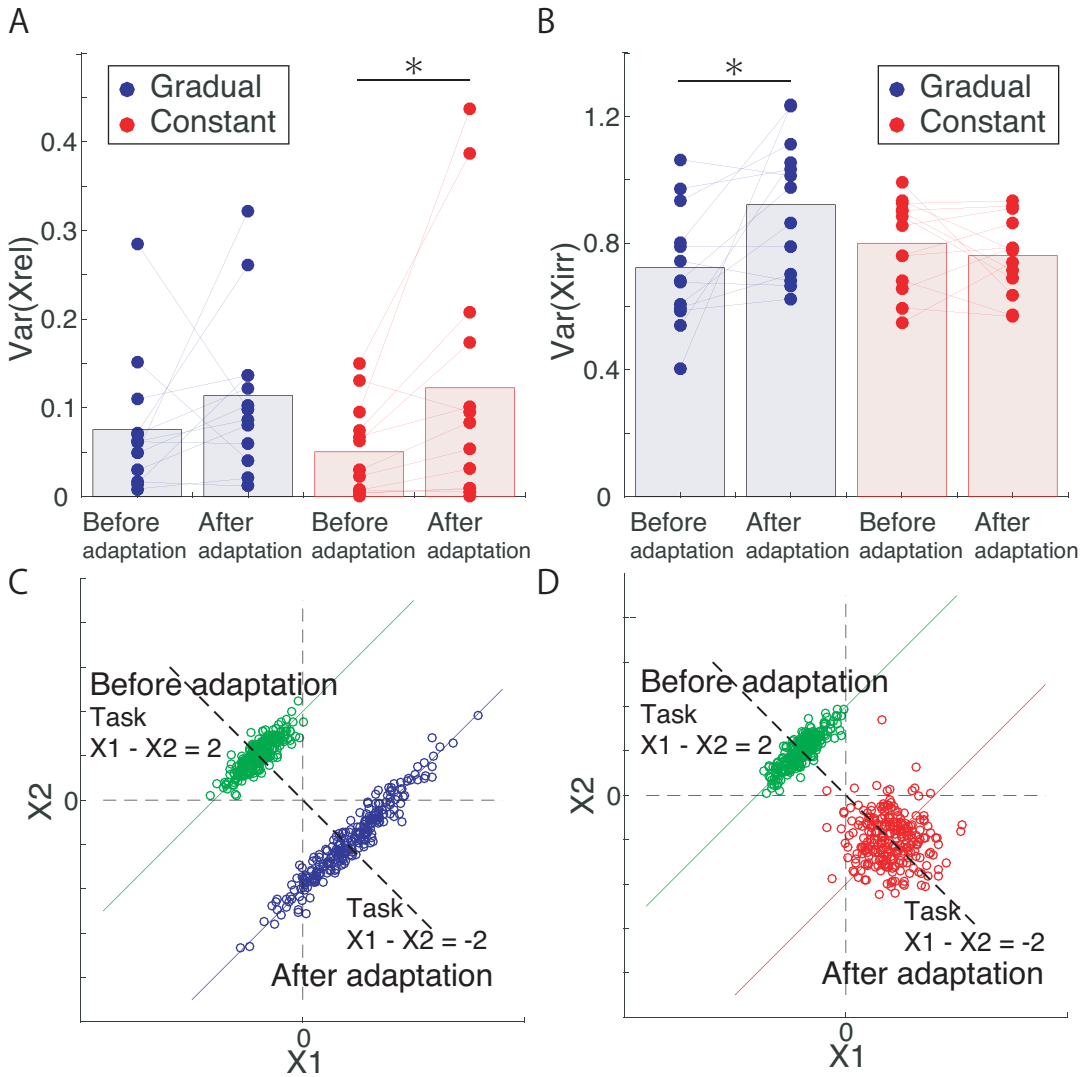
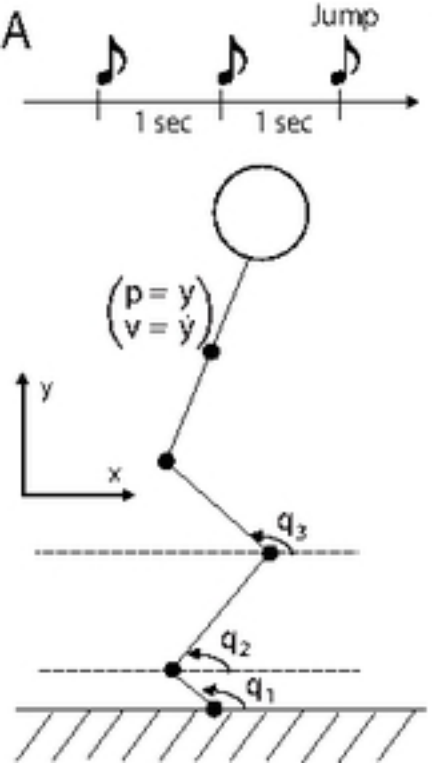
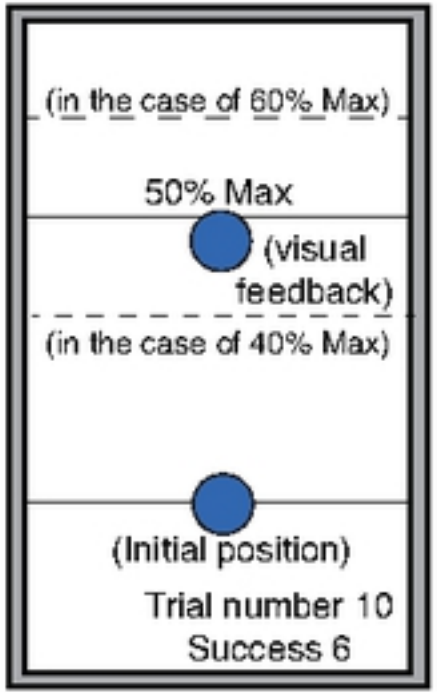
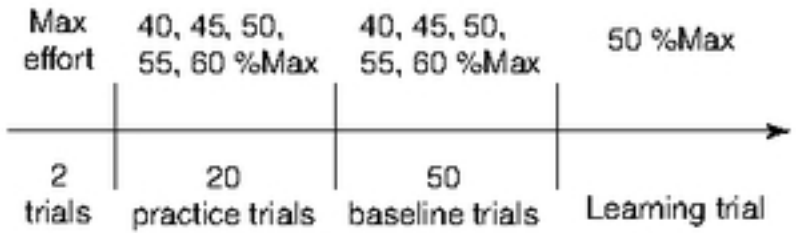


Figure 6. Application of our method to the results of experiments 2 and 3. **A**: Task-relevant variabilities of each subject (N = 13) before and after adaptation to perturbation in experiments 2 and 3. The blue and red dots indicate the variability of each subject in experiments 2 and 3, respectively. The blue and red lines indicate the modulation of the variability due to adaptation. The blue and red bars indicate the averaged variability across all subjects. There was a significant difference between the variabilities before the adaptation and those after the adaptation in experiment 3 (Wilcoxon signed rank test, p = 0.0105). **B**: Task-irrelevant variabilities of each subject (N = 13) before and after adaptation to perturbation in experiments 2 and 3. There was a significant difference between the variabilities before the adaptation and those after the adaptation in experiment 2 (Wilcoxon signed rank test, p = 0.0215). **C,D**: Interpretation of our results based on a simple example. We assume that the task before adaptation required $X_1 - X_2$ to be 2 and that the task after adaptation required $X_1 - X_2$ to be -2. Panel C indicates an interpretation of our results in experiment 2. In the experiment, task-irrelevant variabilities increased after adaptation, and task-relevant variabilities remained unchanged. Panel D indicates an interpretation of our results from experiment 3. In the experiment, task-relevant variabilities increased after adaptation, and task-irrelevant variabilities remained unchanged.

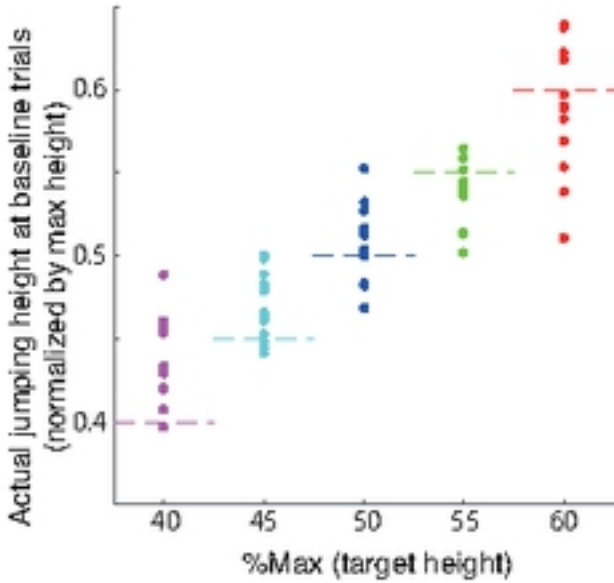


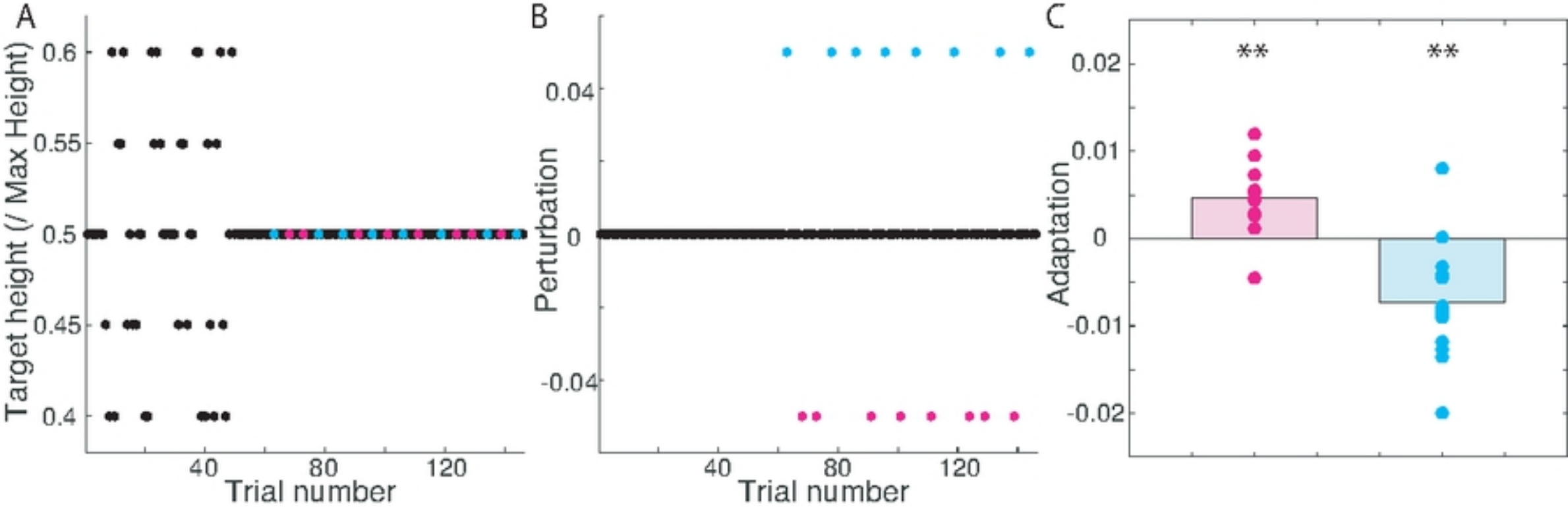


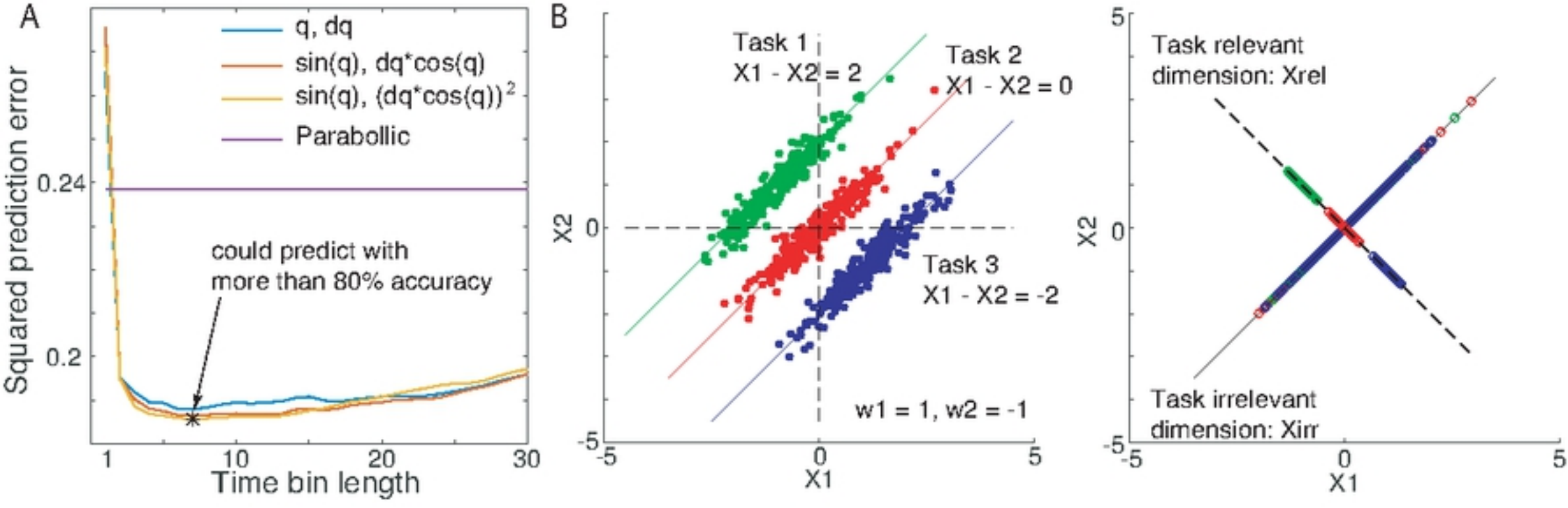
C

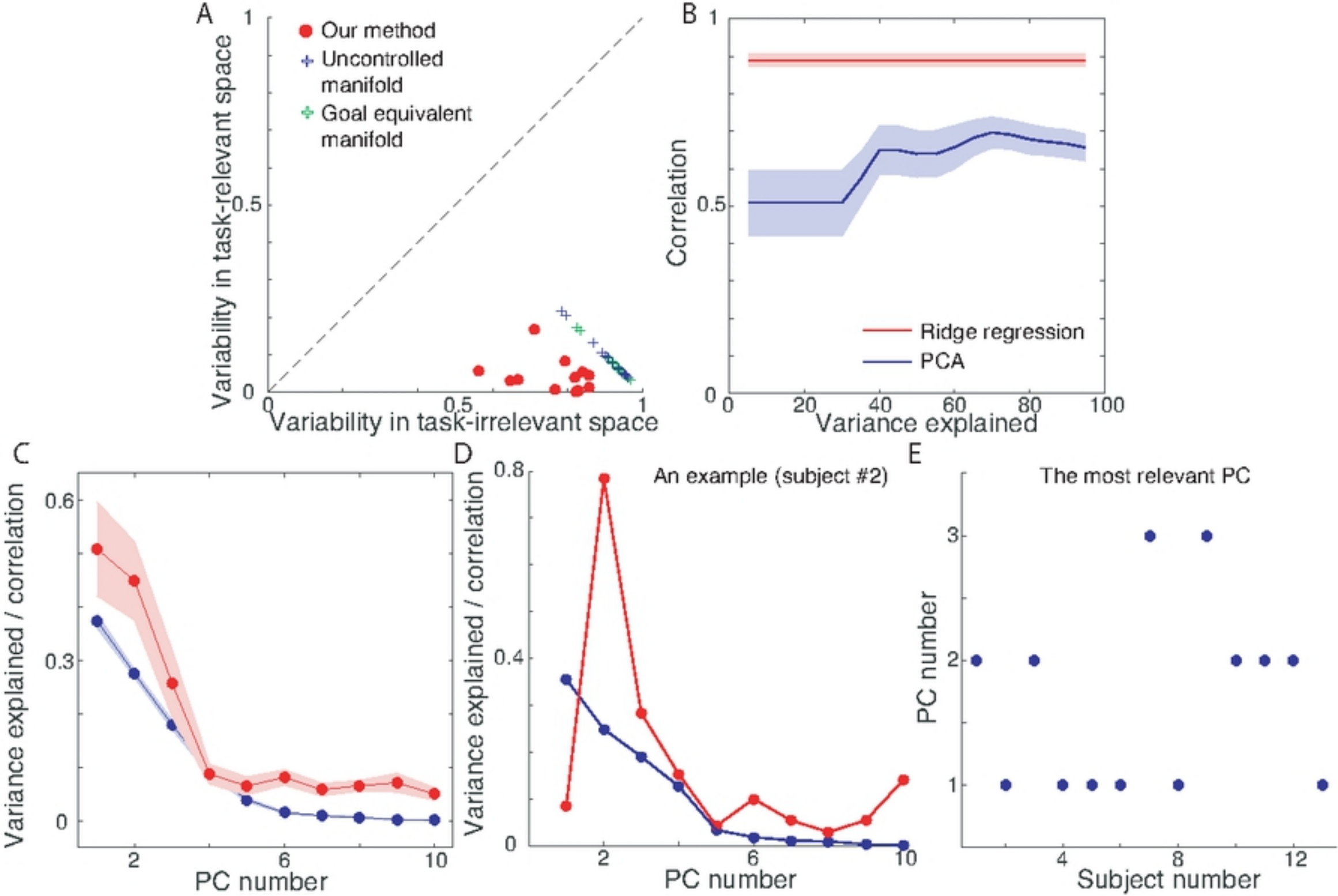


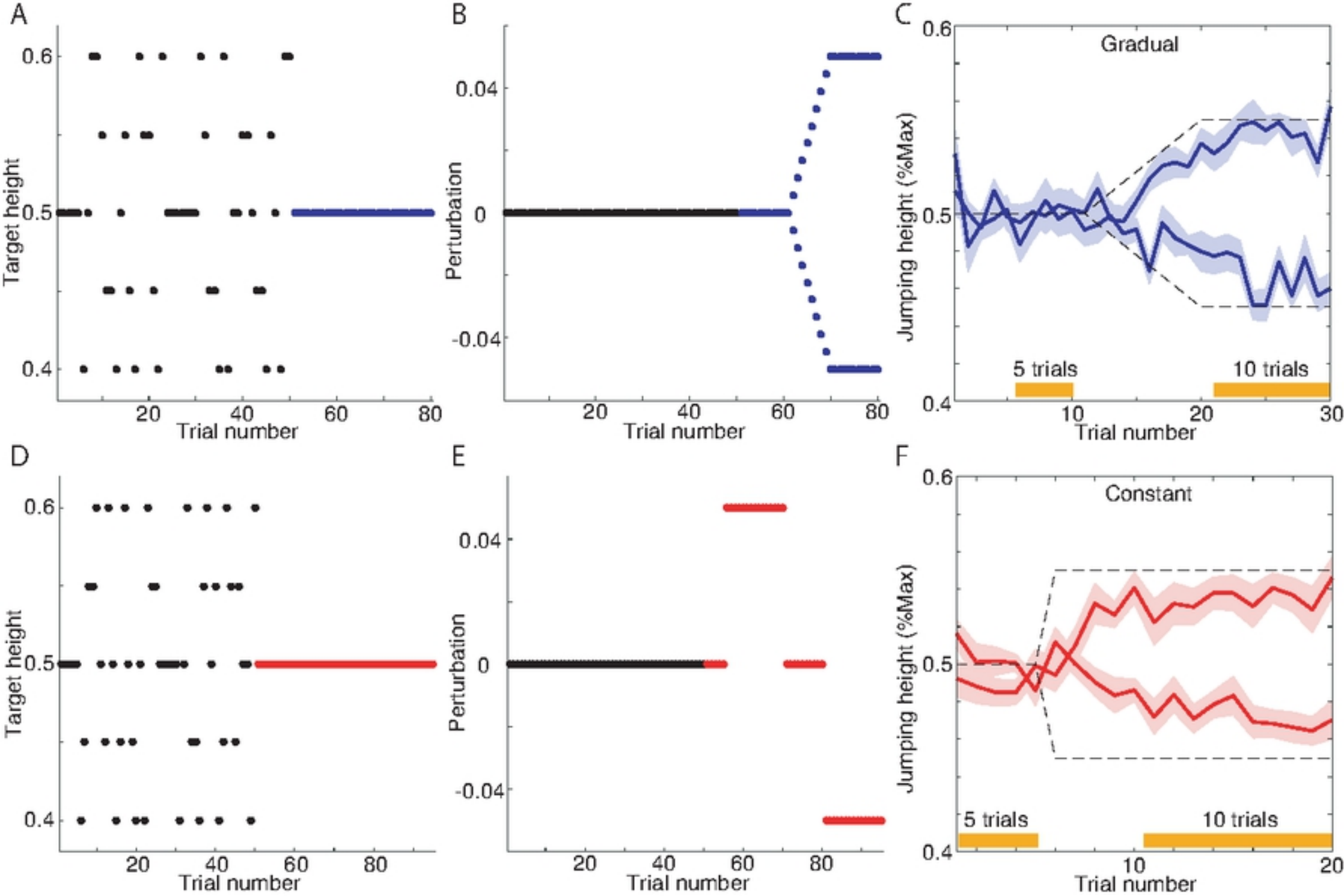
	# of trials	Perturbation
Experiment 1	96	Once in 5 trials
Experiment 2	30 (×2 days)	Gradually increased
Experiment 3	45	Constant

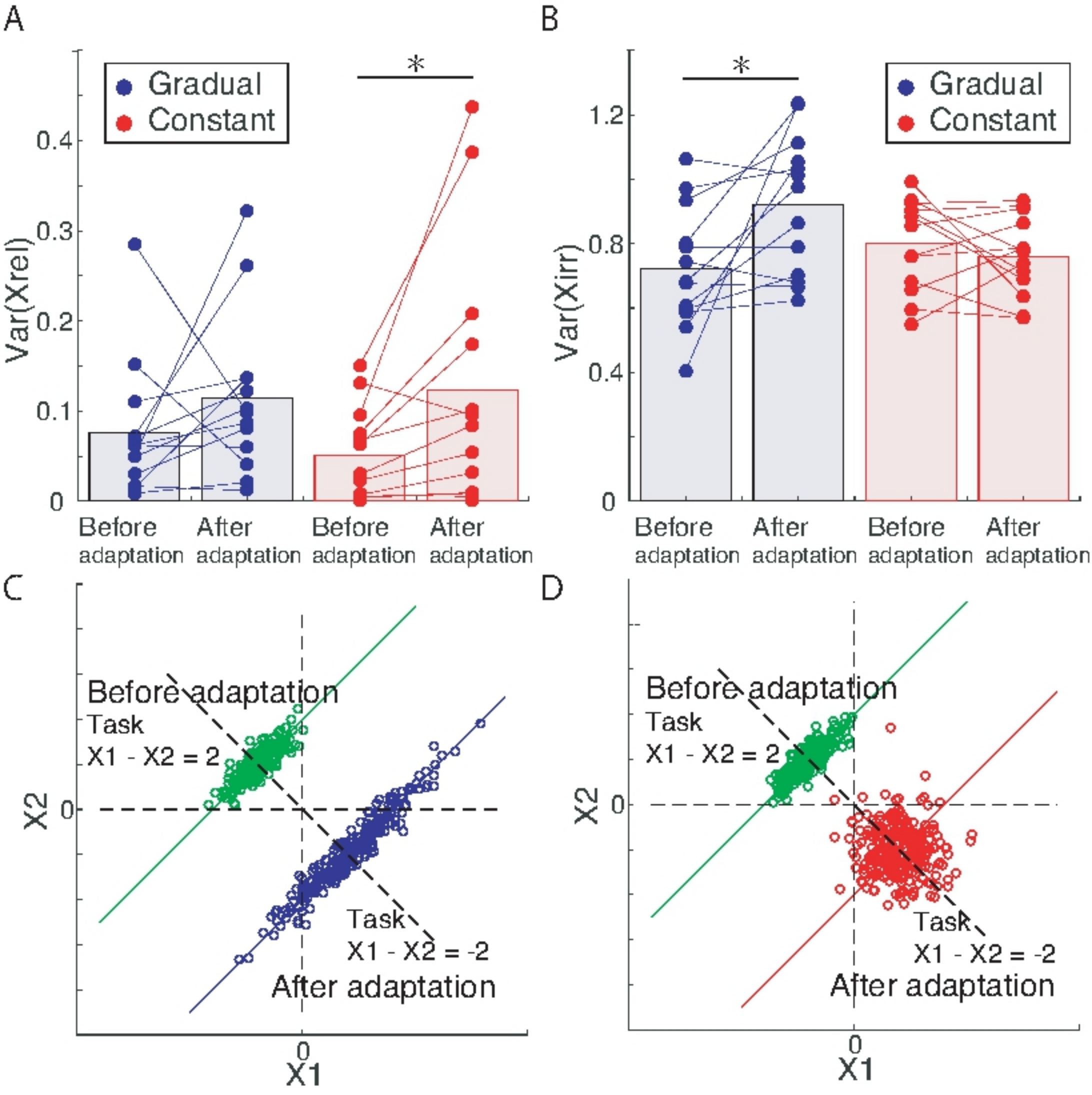



A
Target height (/ Max Height)
0.6
0.55
0.5
0.45
0.4
40
80
120
Trial number
B
Perturbation
0.04
0
-0.04
40
80
120
Trial number
C
Adaptation
0.02
0.01
0
-0.01
-0.02
**
**


A
q, dq
sin(q), dq*cos(q)
sin(q), (dq*cos(q))^2
Parabollic
Squared prediction error
0.24
0.2
could predict with
more than 80% accuracy
1
10
20
30
Time bin length
B
Task 1
X1 - X2 = 2
Task 2
X1 - X2 = 0
Task 3
X1 - X2 = -2
w1 = 1, w2 = -1
X2
X1
5
0
-5
Task relevant
dimension: Xrel
Task irrelevant
dimension: Xirr


A
Our method
Uncontrolled manifold
Goal equivalent manifold
Variability in task-relevant space
Variability in task-irrelevant space
0
0.5
1
B
Correlation
Ridge regression
PCA
Variance explained
0
20
40
60
80
100
C
Variance explained / correlation
PC number
2
6
10
0.6
0.3
D
An example (subject #2)
Variance explained / correlation
PC number
0.8
0.4
E
The most relevant PC
PC number
Subject number
4
8
12
3
2
1


A
B
C
D
E
F
Target height
Perturbation
Jumping height (%Max)
Trial number
Gradual
Constant
5 trials
10 trials
0.6
0.5
0.4
0.04
0
-0.04
20
40
60
80
10
30


A
B
C
D
Gradual
Constant
Var(Xrel)
Var(Xirr)
0
0.1
0.2
0.3
0.4
0.4
0.8
1.2
*
Before adaptation
After adaptation
Before adaptation
Task
X1 - X2 = 2
Task
X1 - X2 = -2
After adaptation
X1
X2