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 is discussed from several various perspectives in motor neuroscience and biomechanics; it can originate from the variability of neural activities, it can reflect a large degree of freedom inherent in our body movements, it can decrease muscle fatigue, or it can facilitate motor learning. How to evaluate motor variability is thus a fundamental question in motor neuroscience and biomechanics. Previous methods have quantified (at least) two striking features of motor variability; the smaller variability in the task-relevant dimension than in the task-irrelevant dimension and the low-dimensional structure that is often referred to as synergy or principal component. However, those previous methods were not only unsuitable for quantifying those features simultaneously but also applicable in some limited conditions (e.g., a 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 the 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. Furthermore, our setting enables the induction of motor adaptation by using perturbation and evaluating the modulation of task-relevant and task-irrelevant variabilities through motor adaptation. Our method enables the identification of a novel property of motor variability; the modulation of those variabilities differs depending on the perturbation schedule. Although a gradually imposed perturbation does not increase both task-relevant and task-irrelevant variabilities, a constant perturbation increases task-relevant variability.

Introduction

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

same movements. Our motor systems somehow tame those 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 an infinite number of joint configurations, muscle activities, and neural activities that can correspond to the desired movement [7-10]. Our motor systems somehow resolve those difficulties (i.e., variability and a large DoF) and generate the desired movements.

Although it remains unclear how we tame movement variability, a possible answer is the decomposition of motor variability into task-relevant and task-irrelevant variabilities. We compensate for a portion of motor variability that is relevant to achieve the 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 the achievement of the desired movements in the presence of movement variability.

Several studies have developed 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 kinematic parameters relevant to task achievements, 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 those kinematic parameters concerning joint angles or angular velocities, enables the definition of the null space around the joint angles or angular velocities averaged across trials. The variability along the null space can be defined as the task-irrelevant variability. Those 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 focused on forward kinematics that map joint angles and angular velocities into joint positions and velocities in the external coordinate. In contrast, tolerance, noise, and covariation analysis (TNC) [13] and goal-equivalent manifold analysis (GEM) [14] focused 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 the gravitational acceleration. When controlling h is a task, the relation among h, p, and v can be task function. For example, with a small value d (i.e., $d^2 \simeq 0$), a slight change in the release position $p + \frac{v}{g}d$ and the release velocity $v - d$ does not cause any change in h; thus, the variability along these slight changes to be regarded as the task-irrelevant variability. TNC and GEM evaluate the task-relevant and task-irrelevant variabilities based on the task functions.

Those techniques have pros and cons. UCM enables the evaluation of motion sequence, but it does not consider the task function. The framework is thus not always suitable for the situation when kinematic parameters are nonlinearly relevant to task achievements, such as the quadratic function of v in the parabolic mentioned above. Because forward kinematics are nonlinear functions of joint angles and angular velocities, UCM requires local linear approximation around representative joint angles or angular velocities based on the Jacobian matrix. Due to the linear approximation, UCM assumed the variability around the kinematics averaged across all the trials. The approximation results in difficulty simultaneously considering the motor variability when the averaged kinematics change, such as those before, during, and after motor learning (although it is possible to discuss those situations separately [18]). GEM also considers the local linear approximation of the nonlinear task function; thus, the method also considers the variability around the task parameters averaged across all the trials. Although GEM can deal with the task function, it is difficult to consider motion sequence in several cases. For example, to consider the motion sequence, the GEM framework needs to define how the dart or ball position and velocity 100 msec before the release are relevant to the maximum height in the example, as mentioned above. TNC enables the simultaneous discussion of the motor variability before, during, and after motor

learning because the framework captures the whole variability in a nonparametric manner without local linear approximation; however, TNC is not always suitable for considering motion sequence based on a similar reason to GEM - it requires an explicit definition of the task function. In total, each method has pros and cons; thus, no single framework can simultaneously evaluate task-relevant and task-irrelevant variabilities when the averaged kinematics or task parameters change while considering both the motion sequence and task function.

Motor variability has another striking feature: the variability is embedded in a 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].

The motor variability thus has at least two characteristics: compensation of the task-relevant variability and low-dimensional structure. Most of the techniques, however, deal with only one aspect. It is difficult to detect the low-dimensional structure of motor variability by the methods to evaluate task-relevant and task-irrelevant variabilities. Similarly, it is difficult to evaluate task-relevant and task-irrelevant variabilities by the techniques to extract low-dimensional structure of motor variability. Thus, the 1st principal component, the dimension that can explain the most significant portion of variability among all dimensions, is not always the most relevant or irrelevant to task performance. Although UCM has been used to extract synergy [8], the primary advantage of UCM is not the extraction of the 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 [26, 27], LDA enables only discrimination, e.g., success or failure of the movement [28], in contrast to TNC and GEM, which can discuss motor performance based on continuous performance value. In summary, few methods simultaneously quantify 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 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 PCA. The current study relied on a ridge regression [29], a linear regression technique that is robust in the presence of measurement noise, has a definite relation to PCA, and can evaluate how the motion of each body part at each time is relevant to task performance in a data-driven manner [30]. Our technique can thus enable the identification of task functions in a data-driven manner without any explicit function such as the parabola or the forward kinematics. First, we formalize the dissociation of motion sequence data into task-relevant components and task-irrelevant components by extending a ridge regression. Second, 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. Third, we validate the decomposition of motion sequence data into task-relevant and task-irrelevant components based on our experimental data. Fourth, we clarify the relation between ridge regression and PCA, a popular method to extract the low-dimensional space of motor variability. 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 to discuss the modulation of movement variability before and after motor adaptation, we discuss the dependence of the modulation on the perturbation schedule.

Results

Downloading our program code on our website is possible.

Linear regression. The current study relied on linear regression to determine the relationship between motion data $\boldsymbol{X} \in \mathrm{R}^{T\times D}$ (the current study focused on the temporal sequence of joint angles and angular velocities) and performance data $\boldsymbol{d} \in \mathrm{R}^{T\times 1}$ following $\boldsymbol{h} = \boldsymbol{X}\boldsymbol{w}$, where T and D denoted the number of trials and the number variables in the motion data, $\boldsymbol{h} \in \mathrm{R}^{T\times 1}$ is the predicted performance, and $\boldsymbol{w} \in \mathrm{R}^{D\times 1}$ is the best linear coefficients to predict the performance [30]. $\boldsymbol{X}_t$, the tth row of $\boldsymbol{X}$ or the motion data at the tth trial, consists of vectorized motion data (e.g., after measuring joint angles of knee $\boldsymbol{q}_{\mathrm{k},t} \in \mathrm{R}^{1\times F}$ and hip $\boldsymbol{q}_{\mathrm{h},t} \in \mathrm{R}^{1\times F}$ for F time frames at the tth trial, $\boldsymbol{X}_t = (\boldsymbol{q}_{\mathrm{k},t}, \boldsymbol{q}_{\mathrm{h},t})$). We relied on a ridge regression to predict performance; the method enabled to predict the performance with higher accuracy than the conventional method as mentioned below. Ridge regression is a linear regression technique that is robust against observation noise, is applicable to data with multicollinearity (see *Materials and Methods* for details). Although we relied on a ridge regression to estimate $\boldsymbol{w}$, the following decomposition of input data into the task-relevant and task-irrelevant components could be applied to any linear regression technique.

Decomposition into task-relevant and task-irrelevant components. After estimating the best linear coefficients $\boldsymbol{w}$ based on measured performance data $\boldsymbol{y} \in \mathrm{R}^{T\times 1}$ and motion data $\boldsymbol{X}$, the estimated $\boldsymbol{w}$ enabled not only the prediction of performance but also the decomposition of motion data into a task-relevant component $\boldsymbol{X}_{\mathrm{rel}}$ and task-irrelevant component $\boldsymbol{X}_{\mathrm{irr}}$. By minimizing the cost function

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

under the constraint $\boldsymbol{X} \neq \boldsymbol{X}_{\mathrm{rel}}$ to avoid the 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 appropriate normalization (i.e., mean and standard deviation of each component of $\boldsymbol{X}$ and $\boldsymbol{y}$ were set to be 0 and 1, respectively, see *Materials and Methods* for details), $\boldsymbol{X}_{\mathrm{rel}}\boldsymbol{w} = \boldsymbol{X}\boldsymbol{w} = \boldsymbol{h}$ and $\boldsymbol{X}_{\mathrm{irr}}\boldsymbol{w} = 0$, indicating that $\boldsymbol{X}_{\mathrm{rel}}$ and $\boldsymbol{X}_{\mathrm{irr}}$ denoted task-relevant and task-irrelevant components under the framework of linear regression. An important point of this framework is that it does not require any explicit function (e.g., forward kinematics such as in UCM or task function such as in GEM and TNC) but require only data $\boldsymbol{X}$ and $\boldsymbol{y}$.

Figs. 1A and 1B demonstrate typical examples of the decomposition when $\boldsymbol{X}$ includes only 2 elements and constrains the task by setting $h = X_1 - X_2$ (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 patterns of $\boldsymbol{X}$ values resulted in an identical h 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. 1B 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 $\boldsymbol{X}_{\mathrm{irr}}$ plotted on the solid line is not separated based on task.

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 the desired movement

by controlling a large number of DoFs. We focus on a simplified version of whole-body movements: a vertical jump while crossing arms in front of the trunk (Fig. 2A). This goal-directed whole-body movement enabled us to focus on lower limb and trunk 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. 2B, see *Materials and Methods* section for details). 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 k_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 [31, 32]. 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. 2C), the target height pseudorandomly changed in each trial. There was a significant main effect of target height in jumping height (Fig. 2D, one-way repeated measure ANOVA, $p = 6.114 \times 10^{-24}$). 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. 2C), 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. 3A and B). We observed the modification of jumping height after each perturbation (Fig. 3C, paired t-test, $p = 0.0026$ for motor adaptation when $p_t = 0.05$; and $p = 0.0014$ for motor adaptation when $p_t = -0.05$). Additionally, we confirmed whether fatigue influenced the adaptation by comparing the magnitudes of the modification in the former part of the learning trials to those in the latter part. There was no significant difference in the magnitudes of the modification between the former and the latter parts of the learning trials (paired t-test, $p = 0.4382$). Motor adaptation could thus be observed in the goal-directed vertical jump without significant effect of fatigue.

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 *Materials and Methods*). 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 [30].

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 [28]. The current study focused on prediction error between actual and predicted jumping height using 10-fold cross-validation. We compared the following three types of input data (see *Materials and 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 $\{a_i\} = (a_1, a_2, a_3, a_4)$ and $\dot{q}_i$ denotes the derivative of q_i concerning time (the definitions of each q_i are given in Fig. 2A). The second candidate is the functions of q_i and $\dot{q}_i$, which describe the position and velocity of back 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 $\{a_i a_j\} = (a_1^2, a_1 a_2, a_1 a_3, a_1 a_4, a_2^2, a_2 a_3, a_2 a_4, a_3^2, a_3 a_4, a_4^2)$. By comparing these three candidates, we found that the first candidate (i.e., $\boldsymbol{X} = (\{q_i\}, \{\dot{q}_i\})$) showed the lowest prediction error (Fig. 4A). In particular, the first candidate, with four-time frames before release, yielded the lowest prediction error. 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. 4A, the first candidate with four-time frames resulted in a prediction error of 0.174, indicating that the ridge regression enables prediction of jumping height with an 82.6±2.28% (mean ± standard error of the mean [s.e.m.], N=13) accuracy in the current setting. In the following, we refer to the first candidate with four-time frames as the motion sequence. In our experimental setting with goal-directed vertical jump, $\boldsymbol{X}$ included 32 elements in each trials (4(dim)×4(time frames) for $\{q_i\}$, and 4(dim)×4(time frames) for $\{\dot{q}_i\}$. 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±2.96% (mean ± s.e.m., N=13) accuracy (purple line in Fig. 4A), 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.

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. The current study calculated the variability (variance) of each element of $\boldsymbol{X}_{\mathrm{rel}}$ and $\boldsymbol{X}_{\mathrm{irr}}$ in focused trials (see *Materials and Methods* for details). 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.

We found that the task-relevant variability was smaller than the task-irrelevant variability in all participants (N = 13, red dots in Fig. 4B). 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 *Materials and 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 properties of 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. Because the low-dimensional structure is considered to represent some features of motor control, it can be expected to be correlated to task performance. 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 *Materials and Methods* for detail). In our setting, there was no clear relation between the number of PCs for the decomposition and the correlation between the decomposed motion data and performance data (Fig. 4C). If averaged across all participants, the 1st PC could explain approximately 30% of the movement variability (blue line in Fig. 4D). Corresponding to the explained movement variability, the 1st PC showed the highest correlation to jumping height (red line in Fig. 4D) 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. 4E). This typical subject was not an exception; Fig. 4F 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 3 out of 13 subjects, the 2nd PC showed the highest correlation, and the 3rd PC showed the highest in 4 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. 4C) 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 *Materials and 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 the 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, 33]. Although there are some differences between adaptation to a constant perturbation and that to a gradually imposed perturbation, e.g., retention rate or awareness [34], 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 was gradually imposed for ten trials and was set to 0.05 or -0.05 for ten trials (Figs. 5A, B). The gradually imposed perturbation required not abrupt but gradual compensation (i.e., subjects were required to modify their motions slightly in each trial). 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). In contrast to experiment 2 where the perturbation was gradually imposed, subjects were required to modify their motions abruptly in experiment 3. Subjects who experienced a $p_t = 0.05$ 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 = 0.1909). 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 = 0.0034). For task-irrelevant variability, there was no significant difference before and after adaptation to gradually increasing or decreasing perturbations (blue dots in Fig. 6B, N = 13, Wilcoxon signed rank test, p = 0.1677) and a constant perturbation (red dots in Fig. 6B, N = 13, Wilcoxon signed rank test, p = 0.3396). These results could be interpreted based on a simulated and two-dimensional case similar to that shown in Fig. 1 (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, there was no modulation in both task-relevant and task-irrelevant variabilities (Fig. 6C). In adapting to a constant perturbation, the task-relevant variability increased, while the task-irrelevant variability was not be modulated (Fig. 6D). Notably, Figs. 6C and 6D were not real data but simulated examples to interpret our results. In summary, the modulation of task-relevant variability 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 (Fig. 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.

Our method can be regarded as a generalized method including UCM and GEM. When we define $X_i = q_i - \bar{q}_i$ (q_i and $\bar{q}_i$ indicated the ith joint angle and the averaged joint angle across all the focused trials, respectively [i=1, ..., 4 in our setting]), and $X_{i+4} = \dot{q}_i - \bar{\dot{q}}_i$ ($\dot{q}_i$ and $\bar{\dot{q}}_i$ indicated the joint angular velocity and averaged joint angular velocity across all the focused trials, respectively), and the corresponding weight $\boldsymbol{w}$ as the Jacobian matrix of forward kinematics $p = p(\boldsymbol{q})$ (the back position) and $v = v(\boldsymbol{q}, \dot{\boldsymbol{q}})$ (the back velocity) around the averaged joint angles and angular velocities across all the focused trials $\bar{\boldsymbol{q}}, \bar{\dot{\boldsymbol{q}}}$, our framework corresponds to UCM. When we define $w_1 = 1, w_2 = \frac{\bar{v}}{g}, X_1 = p - \bar{p}$, and $X_2 = v - \bar{v}$, $h = \boldsymbol{X}\boldsymbol{w} = p - \bar{p} + \frac{\bar{v}}{g}(v - \bar{v})$, our framework corresponds to the GEM in the cases when the task function can be defined by the parabolic function, where g denotes the gravitational acceleration. Because the UCM and GEM can be regarded as a special case of our method, our method can be considered as a generalized version of those methods.

Another advantage of our method is the ability to select appropriate input based on predictive power (Fig. 4A). The predictive power also enables the selection of a proper coordinate to define the task performance. A previous study has demonstrated that the UCM and TNC frameworks are sensitive and insensitive to how to select the coordinate (e.g., either relative or absolute angle), respectively [35]. Our framework is likely sensitive to how to choose the coordinate; however, in contrast to the UCM framework, our method enables the selection of the appropriate coordinate for discussing the relationship between motion and performance based on predictive power. Although we considered one-dimensional performance in the current study (i.e., jumping height), two-dimensional performance requires the definition of an appropriate coordinate to discuss performance [30, 36]. Predictive power plays a vital role in selecting the proper coordinate not only in motion but also in the performance space [30]. How to

select the length of time frames is another crucial problem (Fig. 4A). In our data, the motion data with four-time frames (approximately 33 ms) was chosen for the best predictive power. Although the motion data with four-time frames sound motion fragments rather than motion sequences, our method can be applied independently of the length of time frames. In our case, four-time frames were chosen to increase the predictive power.

Although we compared our method to UCM and GEM (Fig. 4B), we needed to compare it to the TNC method [13,37], the other method used to quantify motor variability from a different perspective. TNC enables the extraction of three types of information from motion data: T-cost, which quantifies how the mean motion data deviate from the optimal motion; N-cost, which quantifies how the motor variability deviates from the optimal variability; C-cost, which quantifies how the covariance among motion data deviates from the optimal covariance. Although the TNC does not consider task-relevant and task-irrelevant variabilities, it can quantify other interesting features (i.e., T-, N-, C-costs) embedded in motion data and variations of those costs during the learning process. Due to the computational cost, we could not apply the TCN method to our data. The TCN requires a grid search in the calculation of T-cost. Because the number of the grid was 200 and that of focused variables was eight in our case (four joint angles and angular velocities), it required 200^8 calculations. Due to this too burdensome computational cost in calculating not only T-cost but also C-cost, we could not apply the TNC method to our case. When the number of focused variables is two, the TNC could be a promising method and work well [13,37].

A potential extension of our method is to yield 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 [38-46]. 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 h_t 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 $\boldsymbol{w}$. Because this is a possible future extension of our approach, we need to further investigate the above-mentioned frameworks.

An advantage of our method is the linearity (i.e., $\boldsymbol{h} = \boldsymbol{X}\boldsymbol{w}$) in contrast to nonlinearity inherent in our body dynamics. A likely explanation for why linear regression works well is by analogy with the motor primitive framework, a successfully used framework in motor adaptation with goal-directed arm-reaching movements [38-46]. In this framework, a nonlinear motor command u is modeled as the linear weighted sum of nonlinear neural activities $\boldsymbol{A}$; $u = \sum_i W_i A_i$ and W_i are modified to minimize the movement error between the actual hand position and the desired movement position. When A_i is a nonlinear function of the desired movement and appropriately high-dimensional, nonlinear motor commands can be generated by appropriate linear combinations of nonlinear neural activities, which has been theoretically validated in the framework of a basis function network [47]. Motion data $\boldsymbol{X}$ can be a nonlinear function of movement performance because our body dynamics are nonlinear. Additionally, the motion data are appropriately high-dimensional (32 dimensions for 1-dimensional performance). Thus, an appropriate linear summation $\boldsymbol{X}\boldsymbol{w}$ could predict the actual movement performance, which resulted in an appropriately estimated $\boldsymbol{w}$ that represented the relevance of motion elements to performance.

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 [28, 48-50], sparse regression technique [51], or nonlinear regression technique [52]. We have shown that a nonlinear regression technique, such as Gaussian process regression, is not effective in predicting performance based on motion data [30], 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. Because some previous studies have discussed the relevance of each PC to performance [53,54], it could be a promising research topic to evaluate the functional roles of low-dimensional structured from a different viewpoint.

To our knowledge, only a few studies have investigated how variability is modulated through motor adaptation [5, 33]. A previous study clarified that the variability is modulated after motor adaptation by utilizing a constant force field [33]. To our knowledge, it was not clarified whether the perturbation schedule affected the modulation. The current study suggests that the perturbation schedule changes the modulation of variability (Figs. 6A and B). Because the variability can facilitate exploration [33], the current study also suggests that constant perturbations facilitate exploration in task-relevant space and that gradually applied perturbations do not affect the exploration. Recent studies have suggested that the variability plays essential roles in sports performance [55], injury prevention [56], and the development of children with developmental coordination disorder [57]. The current result provides a hint about how to assist those functions through the facilitation of the exploration using false feedback.

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 ten 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 heel (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 heel (q_1), right heel and shank (q_2), right shank and thigh (q_3), and right thigh and trunk (q_4) were calculated in the sagittal plane (Fig. 2A). 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 and trunk motions. Throughout the current study, we focused on the four-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. 4A). When the time bin length was four, the fourth time frame corresponded to the release timing, the third 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 display a blue cursor that indicated the height of the marker attached to subject's back and a black bar that indicated target height (Fig. 2B). 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 s in the y-axis (Fig. 2A), k_s, was displayed on the monitor after being normalized for each subject as $k_s = \frac{\hat{k}_s - k_0}{k_{\max} - k_0}$, where $\hat{k}_s$ is the marker position without normalization, k_0 indicates the initial marker position that was evaluated at an upright standing position in each trial, and $k_{\max}$ indicates the jumping height with maximum effort in each subject (Fig. 2C). For the two trials that required the subjects to jump with maximum effort, there was no cursor feedback. One second before the first beep, $k_s = 0$, 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 k_s within each trial (i.e., max k_s), which indicated jumping height (Fig. 2B). When the subjects achieved a jumping motion that was close to the target height ($|d - \max k_s| < 0.02$), 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, the position of the cursor was displayed at max $k_s + p$. 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 ($|d - (\max k_s + p)| < 0.02$), the subjects heard a coin-getting sound to indicate that the jumping motion was successful.

Task-relevant and task-irrelevant variabilities Under the condition $\boldsymbol{X} = \boldsymbol{X}_{\rm rel} + \boldsymbol{X}_{\rm irr}$ (see *Results* for details), the variance of the ith component of $\boldsymbol{X}$, X_i, can be calculated as

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

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

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 [28]. 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. When $\boldsymbol{X}\boldsymbol{X}^T$ has multicollinearity, it is difficult to calculate the inverse of $\boldsymbol{X}\boldsymbol{X}^T$ because of the rank deficit. The identity matrix with a regularization parameter λ enables to calculate the inverse of $\boldsymbol{X}\boldsymbol{X}^T + \lambda\boldsymbol{I}$ and predict output data with a certain accuracy.

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_o, covariance is 0, and the cost function averaged across all the possible noise can be written as

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

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. The equivalence also suggests that the regularization parameter λ corresponds to the variance of the observation noise σ_o^2.

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 $\sigma_i = \sqrt{\frac{1}{T}\sum_{t=1}(X_{i,t}^{\text{original}})^2 - m_d^2}$. 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 σ_i ($\tilde{w}_i = \frac{w_i}{\sigma_i}$) 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.

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 back position p and velocity v of the marker at the release timing as follows

$$h = p + \frac{v^2}{2g}, \tag{8}$$

where $g \simeq 9.8(m/s^2)$. In the joint angle representation, p and v were written as follows:

$$p = \sum_{i=1}^{4} l_i \sin q_i \tag{9}$$

and

$$v = \sum_{i=1}^{4} l_i \dot{q}_i \cos q_i, \tag{10}$$

where l_i indicated the length of the ith limb (i.e., l_1 indicated the length between right toe and heel, l_2 indicated the length between right heel and knee, l_3 indicated the length between right knee and hip, and l_4 indicated the length between hip and back). In the UCM (blue crosses in Fig. 4B), 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 can be written as

$$h = \sum_{i=1}^{4} l_i \sin q_i + \frac{1}{2g}\sum_{i=1}^{4}\sum_{j=1}^{4} l_i l_j \dot{q}_i \dot{q}_j \cos q_i \cos q_j. \tag{11}$$

The first candidate input data for the ridge regression were the joint angles and angular velocities (blue line in Fig. 4A). 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. 4A). The third candidate data were the functions that appeared in equation (11) (orange line in Fig. 4A). In GEM (blue crosses in Fig. 4B), 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 z% explained variance, we determine the number of principal components based on $n_z = \min_n \frac{\sum_{i=1}^{n}\lambda_i^2}{\sum_{i=1}^{\min(T,D)}\lambda_i^2} > \frac{z}{100}$ (i.e., the minimum number of principal components that exceed z% explained variance). After determining n_z, motion data can be reconstructed as $\tilde{\boldsymbol{X}}_t = \sum_{i=1}^{n_z}\lambda_i u_{i,t}\boldsymbol{v}_i$. We then multiply $\sum_{i=1}^{n_z}\boldsymbol{v}_i^T$ by the $\tilde{\boldsymbol{X}}_t$ from the right-hand side, resulting in $\tilde{y}_t = \tilde{\boldsymbol{X}}_t\sum_{i=1}^{n_z}\boldsymbol{v}_i^T = \sum_{i=1}^{n_z}\lambda_i u_{i,t}$. Finally, we calculate the correlation between observed jumping height y_t and $\tilde{y}_t$ in Figs. 4D-4F.

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 interests: The authors declare no competing financial and non-financial interests.

Data availability

All data generated or analyzed during this study are included in this article.

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. *Nat. Rev. Neurosci.* **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. *J. Neurophysiol.* **88**, 1533-1544 (2002).
5. Dhawale, A. K., Smith, M. A. & Ölveczky, B. P. The Role of Variability in Motor Learning. *Ann. Rev. Neurosci.* **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. *J. Neurophysiol.* **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. *Exerc. Sport. Sci. Rev.* **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 Comput. Biol.* **8**, e1002348 (2012).
11. Scholz, J. P. & Sch ö ner, G. The uncontrolled manifold concept: identifying control variables for a functional task. *Exp. Brain Res.* **126**, 289-306 (1999).
12. Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nat. Neurosci.* **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. *Biol. Cybern.* **89**, 22-33 (2003).
14. Cusumano, J. P. & Cesari, P. Body-goal variability mapping in an aiming task. *Biol. Cybern.* **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. *Nat. Neurosci.* **17**, 440-448 (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. *Exp. Brain Res.* **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. *J. Neurophysiol.* **102**, 59-68 (2009).
18. Yang, J.-F., Scholz, J. P. & Latash, M. L. The role of kinematic redundancy in adaptation of reaching. *Exp. Brain Res.* **176**, 54-69 (2006).
19. d'Avella, A., Saltiel, P. & Bizzi, E. Combinations of muscle synergies in the construction of a natural motor behavior. *Nat. Neurosci.* **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. *J. Neurosci.* **27**, 11149-11161 (2007).
21. Hagio, S. & Kouzaki, M. The flexible recruitment of muscle synergies depends on the required force-generating capability. *J. Neurophysiol.* **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. *Sci. Rep.* **5**, 13360 (2015).
23. Lacquaniti, F. & Maioli, C. Independent control of limb position and contact forces in cat posture. *J. Neurophysiol.* **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. Santello, M. & Soechting, J. F. Gradual molding of the hand to object contours. *J. Neurophysiol.* **79**, 1307-1320 (1998).
27. Delis, I., Berret, B., Pozzo, T. & Panzeri, S. Quantitative evaluation of muscle synergy models: a single-trial task decoding approach. *Front. Comput. Neurosci.* **7**(8), (2013).
28. Bishop, C. M. Pattern Recognition and Machine Learning (Springer Verlag, 2006).
29. Hoerl, A. E. & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55-67 (1970).
30. Furuki, D. & Takiyama, K. Detecting the relevance to performance of whole-body movements. *Sci. Rep.* **7**, 15659 (2017).
31. Krakauer, J. W., Pine, Z. M., Ghilardi, M. F. & Ghez, C. Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J. Neurosci.* **20**, 8916-8924 (2000).
32. Cotti, J. et al. Adaptation of reactive and voluntary saccades: different patterns of adaptation revealed in the antisaccade task. *J. Physiol.* **587**, 127-138 (2009).
33. Wu, H. G., Miyamoto, Y. R., Castro, L. N. G., Iveczky, B. P. O. & Smith, M. A. Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat. Neurosci.* **17** 1-13 (2014).
34. Kagerer, F. A., Contreras-Vidal, J. L. & Stelmach, G. E. Adaptation to gradual as compared with sudden visuo-motor distortions. *Exp. Brain Res.* **115**, 557-561 (1997).
35. Sternad, D., Park, S.-W., Mueller, H. & Hogan, N. Coordinate Dependence of Variability Analysis. *PLoS Comput. Biol.* **6**(4):e1000751 (2010).
36. Shinya, M., Tsuchiya, S., Yamada, Y., Nakazawa, K., Kudo, K. & Oda, S. Pitching form determines probabilistic structure of errors in pitch location. *J. Sports Sci.* **35**, 2142-2147 (2017).
37. Cohen, R. G. & Sternad, D. Variability in motor learning: relocating, channeling and reducing noise. *Exp. Brain Res.* **193**, 69-83 (2009).
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. *J. Neurophysiol.* **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. *J. Neurosci.* **23**, 9032-9045 (2003).
41. Smith, M. A., Ghazizadeh, A. & Shadmehr, R. Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS Biol.* **4**, e179 (2006).
42. Takiyama, K., Hirashima, M. & Nozaki, D. Prospective errors determine motor learning. *Nat. Comm.* **6**, 5925 (2015).
43. Takiyama, K. Context-dependent memory decay is evidence of effort minimization in motor learning: a computational study. *Front. Comput. Neurosci.* **9**, 1-10 (2015).

652 44. Takiyama, K. & Sakai, Y. Balanced motor primitive can explain generalization of motor learning
653 effects between unimanual and bimanual movements. *Sci. Rep.* **6**, 23331, (2016).
654 45. Takiyama, K. & Sakai, Y. A balanced motor primitive framework can simultaneously explain motor
655 learning in unimanual and bimanual movements. *Neural Netw.* **86**, 80-89 (2017).
656 46. Ishii, K., Hayashi, T. & Takiyama, K. Influence of switching rule on motor learning. *Sci. Rep.* **8**,
657 13559 (2018).
658 47. Pouget, A., & Sejnowski, T. J. Spatial transformations in the parietal cortex using basis functions.
659 *J. Cogn. Neurosci.* **9**, 222-237 (1997).
660 48. Takiyama, K., Katahira, K. & Okada, M. Exact inference in discontinuous firing rate estimation
661 using belief propagation. *J. Phys. Soc. Jpn.* **78**, 4003 (2009).
662 49. Takiyama, K. & Okada, M. Detection of hidden structures in nonstationary spike trains. *Neural*
663 *Comput.* **23**, 1205-1233 (2011).
664 50. Naruse, Y., Takiyama, K., Okada, M. & Umehara, H. Statistical method for detecting phase shifts
665 in alpha rhythm from human electroencephalogram data. *Phys. Rev. E* **87**, 042708 (2013).
666 51. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Series B.*, 267-288
667 (1996).
668 52. Rasmussen, C. E. & Williams, C. K. I. Gaussian Processes for Machine Learning (MIT Press, 2006).
669 53. Pacheco, M. M. & Newell, K. M. Learning a specific, individual and generalizable coordination
670 function: evaluating the variability of practice hypothesis in motor learning. *Exp. Brain Res.* 1-12
671 (2018).
672 54. Pacheco, M. M. & Newell, K. M. Transfer of a learned coordination function: Specific, individual
673 and generalizable. *Hum. Mov. Sci.* **59**, 66-80 (2018).
674 55. Bartlett, R., Wheat, J. & Robins, M. Is movement variability important for sports biomechanists?
675 *Sports Biomech.* **6**, 224-243 (2007).
676 56. Srinivasan, D. & Mathiassen, S. E. Motor variability in occupational health and performance. *Clin.*
677 *Biomech.* **27**, 979-993 (2012).
678 57. Golenia, L., Bongers, R. M., van Hoorn, J. F., Otten, E., Mouton, L. J. & Schoemaker, M. M.,
679 Variability in coordination patterns in children with developmental coordination disorder (DCD). *Hum.*
680 *Mov. Sci.* **60**, 202-213 (2018).

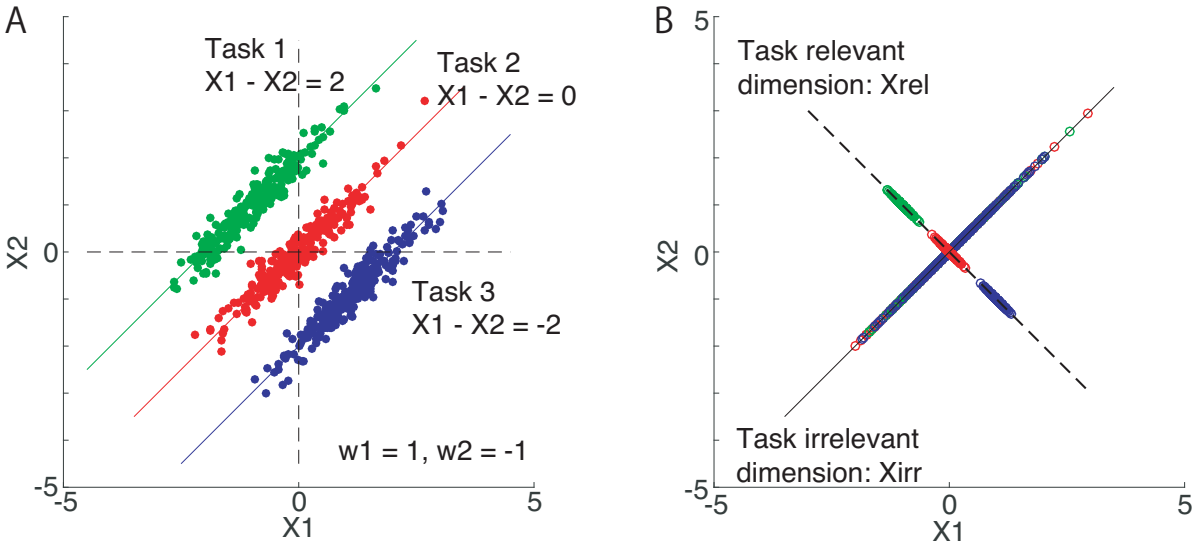

Figure 1

Figure legends

Figure 1. The concept of our method. **A**: An example of decomposing input data $\boldsymbol{X}$ into task-relevant $\boldsymbol{X}_{\mathrm{rel}}$ and task-irrelevant components $\boldsymbol{X}_{\mathrm{irr}}$. 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 (red line), and task 3 required $X_1 - X_2$ to be -2 (blue 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., $h = w_1 X_1 + w_2 X_2 = X_1 - X_2$ should be determined differently in each task. **B**: The input data were decomposed into a task-relevant (black dotted line) component $\boldsymbol{X}_{\mathrm{rel}} = \boldsymbol{X}\boldsymbol{w}\boldsymbol{w}^T/|\boldsymbol{w}|^2$ and a task-irrelevant component $\boldsymbol{X}_{\mathrm{irr}} = \boldsymbol{X} - \boldsymbol{X}_{\mathrm{rel}}$ (solid black line). $\boldsymbol{X}_{\mathrm{rel}}$ was separated depending on the task, and $\boldsymbol{X}_{\mathrm{irr}}$ was not separated, which indicates that the decomposition enables the discussion of the task-relevant and task-irrelevant components.

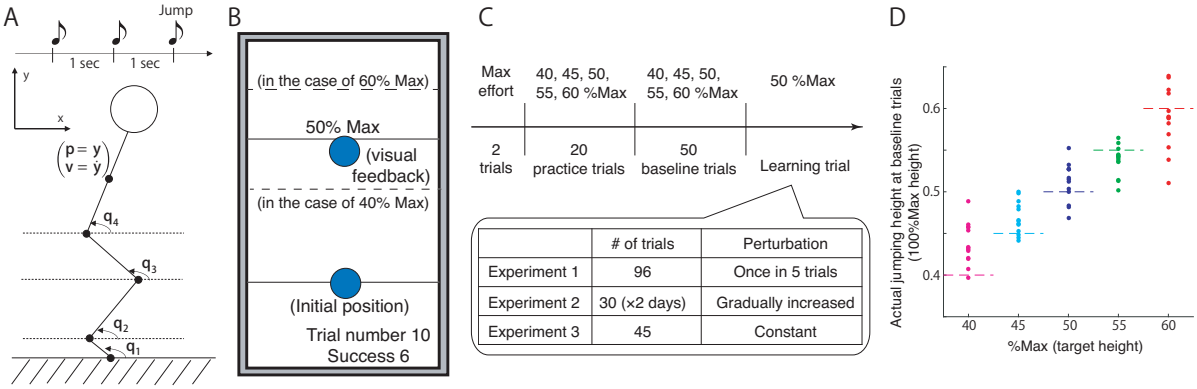

Figure 2

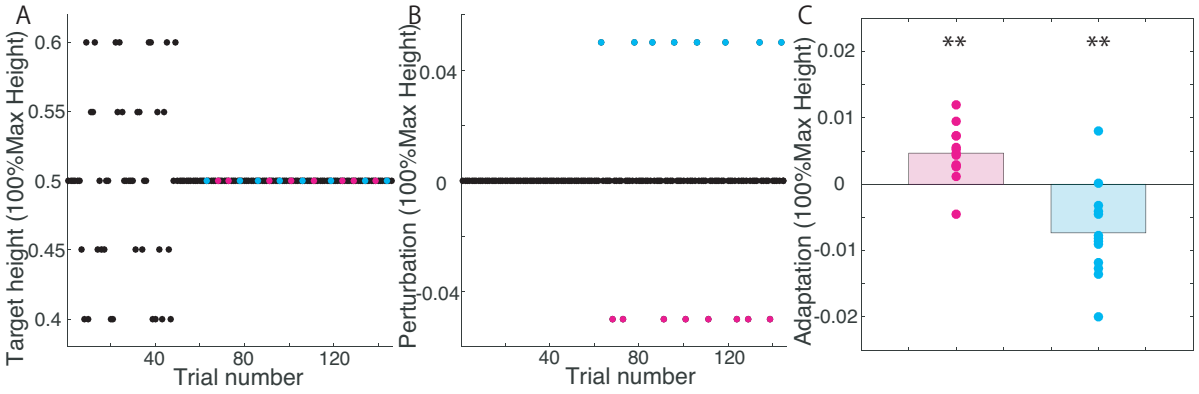

Figure 3

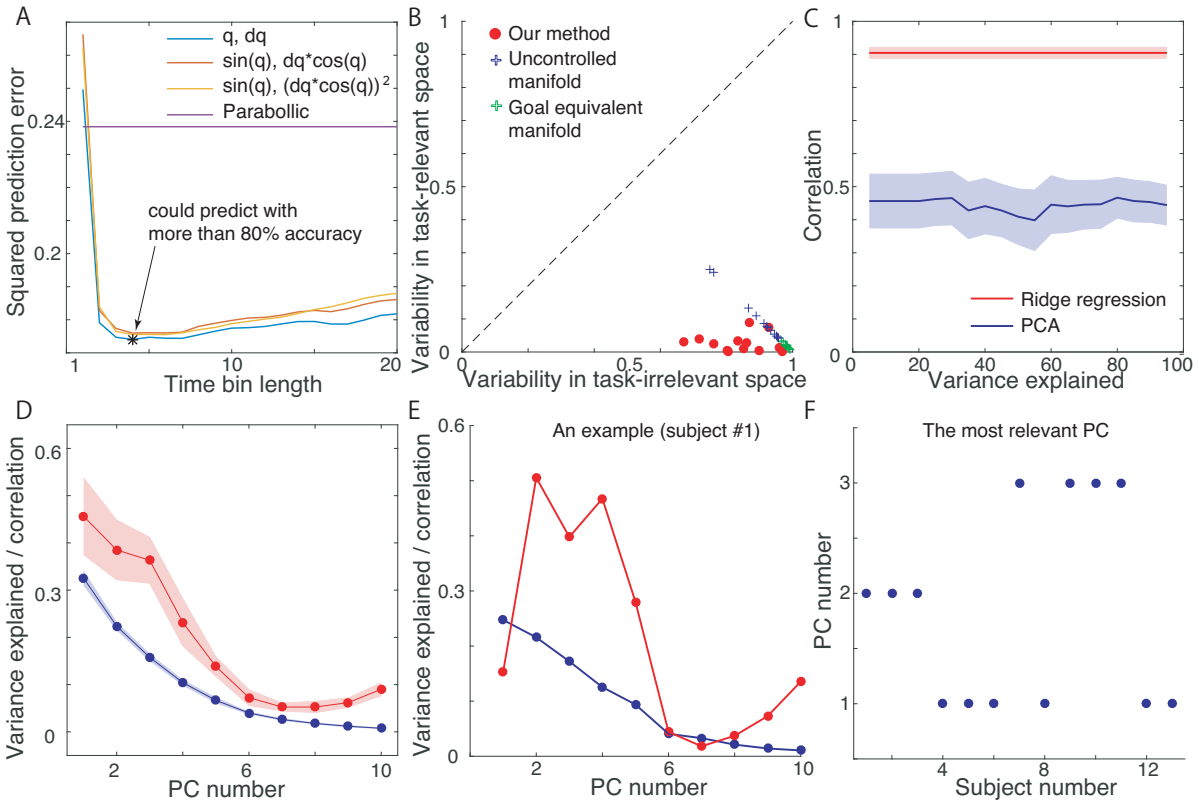

Figure 4

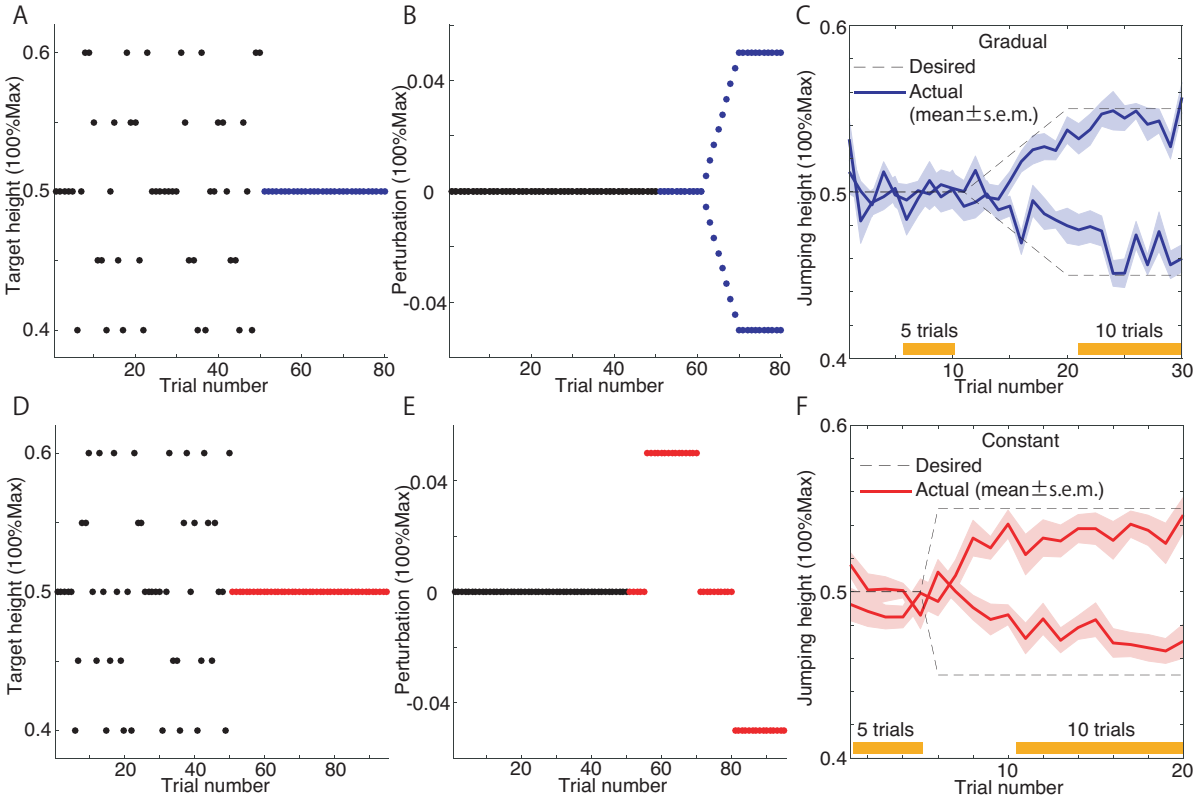

Figure 5

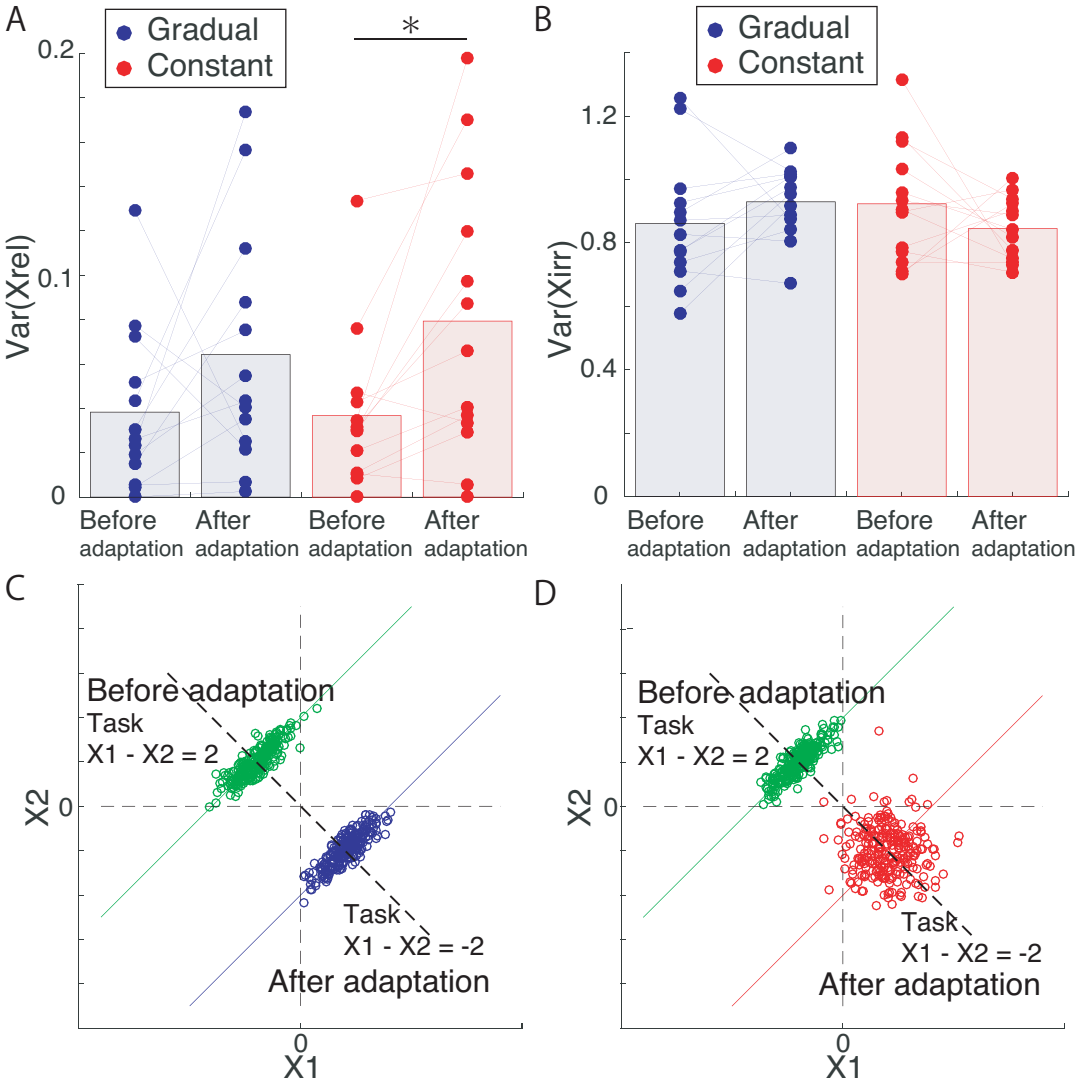

Figure 6

Figure 2. 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, knee, and hip 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 (shown by blue cursor located on the baseline height) were displayed. When the target height was 60%, the black bar and text were presented at the position of the higher black dotted bar. When the target height was 40%, the black bar and text were displayed at the location 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 presented 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. **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 (one-way repeated measure ANOVA, $p = 6.114\times10^{-24}$), indicating that the participants could perform the goal-directed movement.

Figure 3. 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 4. Validation of our method and comparison to previous methods. **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 ridge regression enables the prediction of output data with an 82.6±2.28% (mean ± s.e.m., N=13) accuracy. **B**: 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. **C**: 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. **D**: 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. **E**: 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. **F**: 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 showed 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 show 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 = 0.0034). **B**: Task-irrelevant variabilities of each subject (N = 13) before and after adaptation to perturbation in experiments 2 and 3. **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 experiment 2, there was no modulation in both task-relevant and task-irrelevant variabilities. Panel D suggests an explanation of our findings from experiment 3. In the experiment, task-relevant variabilities increased after adaptation, and task-irrelevant variabilities remained unchanged.