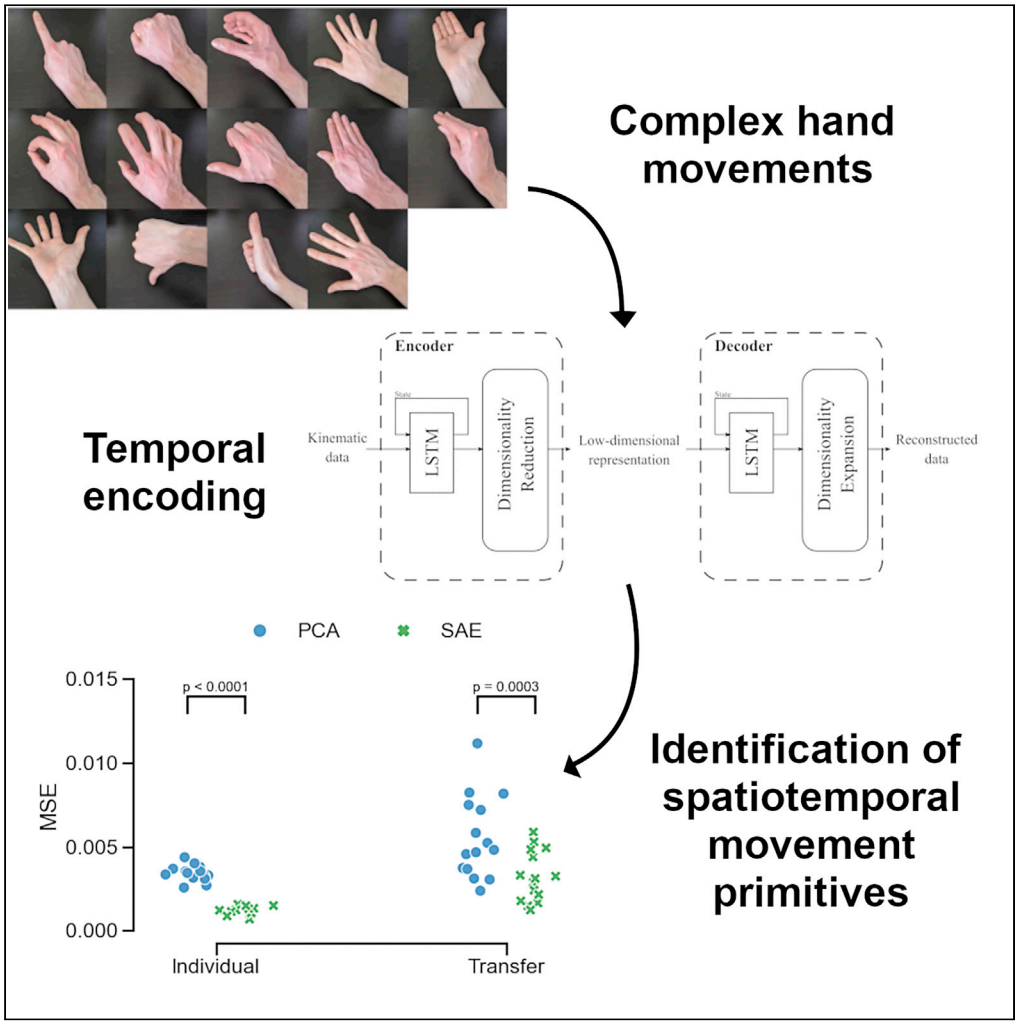
iScience

CellPress
OPEN ACCESS

Article

Decomposition into dynamic features reveals a conserved temporal structure in hand kinematics



Conor Keogh, James J. FitzGerald

james.fitzgerald@nds.ox.ac.uk

Highlights

Hand movements are comprised of a low-dimensional set of movement primitives

Primitive movements have an important temporal component

Spatiotemporal movement primitives are conserved across individuals

New complex movements can be flexibly reconstructed using these primitives

Keogh & FitzGerald, iScience *25*, 105428
November 18, 2022 © 2022
https://doi.org/10.1016/j.isci.2022.105428

iScience

CellPress
OPEN ACCESS

Article

Decomposition into dynamic features reveals a conserved temporal structure in hand kinematics

Conor Keogh[1] and James J. FitzGerald[1,2,*]

SUMMARY

The human hand is a unique and highly complex effector. The ability to describe hand kinematics with a small number of features suggests that complex hand movements are composed of combinations of simpler movements. This would greatly simplify the neural control of hand movements. If such movement primitives exist, a dimensionality reduction approach designed to exploit these features should outperform existing methods. We developed a deep neural network to capture the temporal dynamics of movements and demonstrate that the features learned allow accurate representation of functional hand movements using lower-dimensional representations than previously reported. We show that these temporal features are highly conserved across individuals and can interpolate previously unseen movements, indicating that they capture the intrinsic structure of hand movements. These results indicate that functional hand movements are defined by a low-dimensional basis set of movement primitives with important temporal dynamics and that these features are common across individuals.

INTRODUCTION

The hand is the major effector with which we interact with the world around us. Its complex arrangement of bones, joints, and muscles facilitates the dexterous movements that characterize manual behavior.[1,2] The neural control of the hand is achieved through extremely dense innervation[3] and disproportionate cortical representation[4] with specialized hand motor regions.[5] Despite this neuroanatomical understanding, exactly how the human nervous system accurately controls hand movement remains an open question.[6]

Multiple lines of evidence suggest that the brain does not independently control and co-ordinate each of the many degrees of freedom of the hand. Rather, hand control is greatly simplified by relying on a set of synergies or hand movement primitives.[7–10] Each primitive movement involves the coordinated activity of multiple muscles; these primitives can then be combined in order to generate more complex movements. In this way, hand movements are represented on a low-dimensional manifold of movement primitives,[11–14] allowing operation of an effector with many degrees of freedom with a much simpler control system.[15,16]

In line with this, principal-component analysis (PCA)[17] of hand biomechanics demonstrates that most task-relevant information is contained in the first few principal components.[18–20] It has been shown that single cells in the primary motor cortex produce coordinated activity of multiple muscles[21–24] via spinal interneuron circuits[25] and that direct corticomotoneuronal connections also produce complex synergies of muscle co-activations,[5,26–30] providing a potential neurobiological basis for the apparent intrinsic structure of hand movements.

Understanding the nature and mechanisms of motor synergies is important because it brings us closer to developing methods for the restoration of naturalistic motor function. Although PCA has been successful in identifying low-dimensional features from kinematic data, it is limited to linear combinations of the recorded data. The true motor synergies are likely to be complex functions of the recorded data, in which case dimensionality reduction methods that capture these nonlinear relationships might be expected to outperform PCA. However, recent evidence has demonstrated that using more complex nonlinear dimensionality reduction methods does not by itself better capture the structure of hand movements,[31,32] indicating that the structure seen in low-variance components is not merely due to a failure of PCA to

[1]Oxford Neural Interfacing, Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK

[2]Lead contact

*Correspondence: james.fitzgerald@nds.ox.ac.uk

https://doi.org/10.1016/j.isci.2022.105428

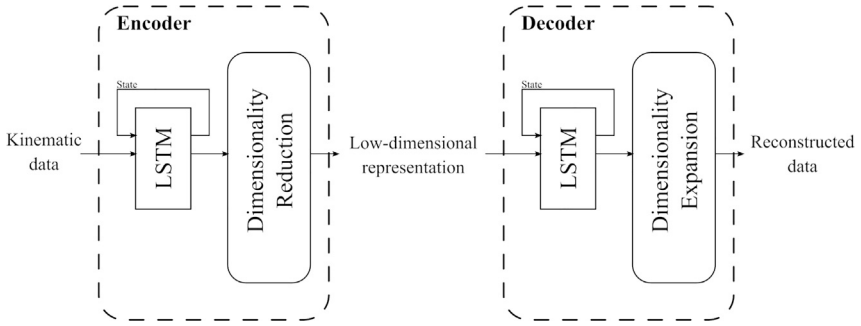


Figure 1. Sequential autoencoder architecture

The network is designed to learn a limited set of features with temporal dynamics that allow maximal reconstruction of the original data. The encoder takes in raw kinematic data and maps this to a set of dynamical features. The encoder is composed of a long short-term memory (LSTM) recurrent neural network, which tracks the state of the network between samples and allows the learning of temporal features, and a dimensionality reduction layer that maps the LSTM output down to a low-dimensional representation. The decoder is constructed similarly. This takes in low-dimensional representations of kinematic data and passes this through an LSTM before expanding the dimensionality to reconstruct the original data from the limited set of features.

capture complex relationships.[31] This has been interpreted as suggesting that functional hand movements may be higher dimensional than previously thought.

Previous work on the neural basis of hand biomechanics has demonstrated a low-dimensional structure of the neural representations of hand movements.[33] These neural synergies are not constant, but vary over time.[14,33] This produces neural synergies with important temporal features.[10,16] Despite the recognized importance of temporal information in neural data, the methods used in previous studies of hand biomechanics do not incorporate the temporal evolution of hand movements. We hypothesized that by developing a method to specifically account for temporal dynamics, we would be able to better learn the underlying latent movement features. We designed a sequential autoencoder to achieve this (Figure 1). This is a deep neural network based on autoencoders, with recurrent networks[34] to learn temporal dynamics in the encoder and decoder stages. Similar approaches have shown promise in analysis of time series of neural spiking data.[35]

We applied this sequential autoencoder to hand kinematic data in order to investigate whether a low-dimensional structure with important temporal dynamics exists in these data, in line with the evidence on the neural mechanisms underlying these movements. We demonstrate that incorporation of temporal dynamics allows learning of features that better capture the structure of hand movements. Using even very low-dimensional representations, highly accurate movement reconstructions could be achieved, significantly outperforming standard linear models.

We show that the learned movement features, representing hand synergies with temporal dynamics, are valid across a population, that population-level models can be applied to new individuals and maintain accuracy, and that these models can reconstruct previously unseen functional movements by recombination of learned features even with a very limited training set, analogous to the proposed mechanisms underlying human motor control. These results indicate that human functional hand movements are low dimensional, defined by a basis set of movement synergies with key temporal dynamic properties. This has important implications for biomechanics and the development of systems for the restoration of functional hand movements.

RESULTS

Functional hand movements possess a latent structure

Fifteen healthy, right-handed volunteers were recruited (9 male, 6 female). All participants carried out 900 trials of functional hand movements while detailed kinematic data were recorded, including hand position, orientation, and acceleration, wrist and thumb angles, and bend at the metacarpophalangeal and distal interphalangeal joints of each finger. Participants performed 14 functional movements, made up of a combination of common hand grips and hand positions. The hand movements used are listed in Table S1. Images of the hand movements used are shown in Figure 2A.

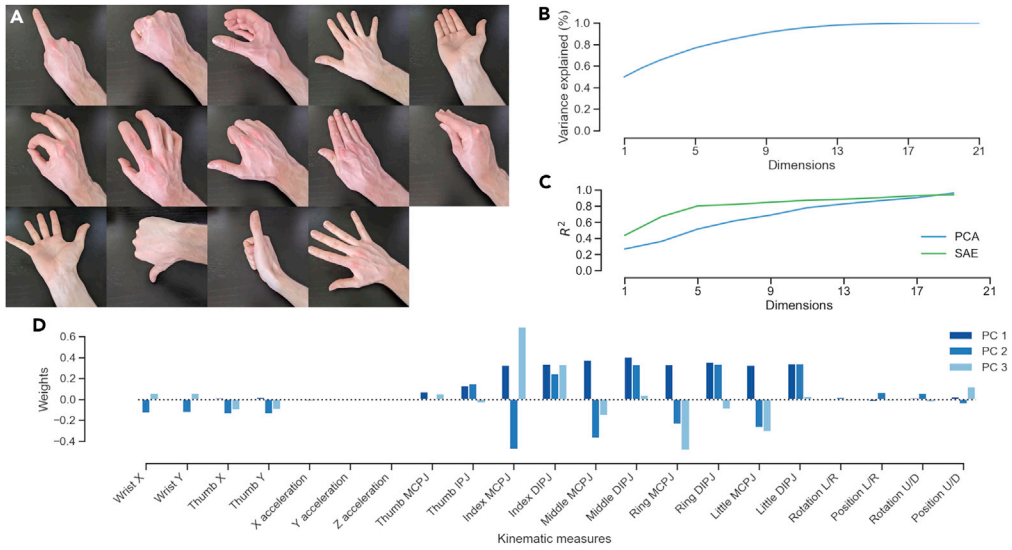


Figure 2. Natural movements contain a latent structure

(A) Functional hand movements used, including a combination of common grips and hand positions.

(B) Variance explained by PCA using a range of dimensions; 9 dimensions account for >90% of the variance in the kinematic data, in line with published results.

(C) Coefficient of determination (R^2) of PCA and SAE models using a range of dimensions; SAE models explain more of the variance in the data with fewer dimensions, with the models converging as all data are included.

(D) Weights of the first three principal components, accounting for most of the variance in the data; this demonstrates some of the structure in the kinematic data that is exploited by dimensionality reduction models. SAE: sequential autoencoder; PC: principal component; MCPJ: metacarpophalangeal joint; IPJ: interphalangeal joint; DIPJ: distal interphalangeal joint; L/R: left/right; U/D: up/down.

Latent features that capture the underlying structure of these natural hand movements can be derived from the data using PCA. The cumulative variance explained by a given number of learned features in a pooled analysis of all participants using PCA is shown in Figure 2B. This demonstrates that 91% of the variance in the data can be accounted for by 9 features using PCA, in agreement with previous studies.[13,36] For this reason, further analyses were performed using 9 features unless stated otherwise, representing a >50% reduction in the data required to encode movements.

Similarly, our sequential autoencoder model is able to learn low-dimensional representations that capture most of the variation in the kinematic data. Figure 2C shows the coefficient of determination (R^2) for both PCA and sequential autoencoder models trained with a range of sizes. The sequential autoencoder model captures more of the variation in the kinematic data with even very low-dimensional representations, demonstrating a superior ability to capture the structure of hand movements. Both approaches converge toward capturing all of the variance in the data as the size of the representation approaches the number of dimensions in the full kinematic dataset.

The weights of the first three principal components across each of the movements are shown in Figure 2D. This demonstrates the structure that is learned by the linear model. Notably, the first principal component is high when all joints are bent, providing a measure of the extent of "grasping," whereas the second principal component appears to measure a sum of metacarpophalangeal joint flexion across all fingers. The third principal component is high when the index finger is differentiated from all other fingers, indicating the importance of the index finger in functional movements.

Temporal dynamics improve movement reconstructions

Detailed hand kinematic data can be accurately reconstructed from small numbers of these latent features. The results of model comparison for reconstruction error using a two-way repeated measures ANOVA with model type (PCA and sequential autoencoder) and data source (individualized models, trained on a subset of the subjects' own data, and transferred models, trained using only data from other subjects and assessed on the subject's own data to simulate the application of a population model to a new individual) as

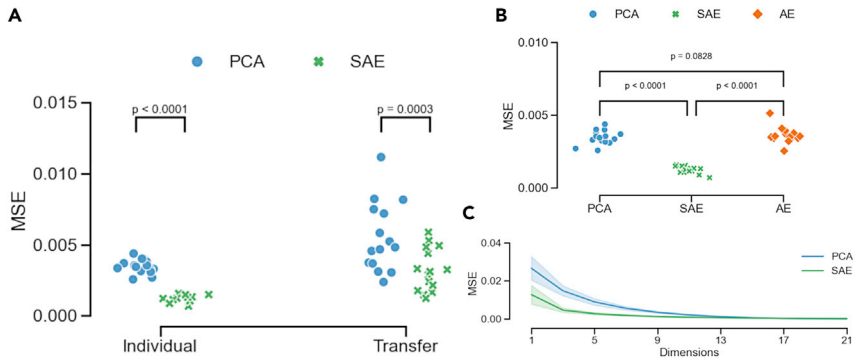


Figure 3. Temporal dynamics improve the reconstruction of hand kinematics from low-dimensional feature spaces

(A) Mean squared error (MSE) reconstruction loss for principal component analysis (PCA) and sequential autoencoder (SAE) models using individualized and transferred data sources. Each point represents the MSE loss for each model type for one participant. The SAE performs significantly better at reconstructing movements. Transferred models perform worse than individualized models, but this effect is less pronounced for the SAE. The performance of transferred SAE models is equivalent to individualized PCA models.

(B) MSE loss for PCA, SAE, and a standard autoencoder (AE) without temporal information. The AE performs equivalently to PCA, whereas the addition of temporal dynamics in the SAE results in improved performance over both. The ability of the SAE model to accurately reconstruct hand kinematics is due to the use of temporal information.

(C) Reconstruction error for models trained with a range of dimensionalities. Increasing the dimensionality of the representation improves reconstruction for both models. The SAE demonstrates superior reconstruction accuracy even with very low-dimensional representation. A 5-dimensional SAE model outperforms a 9-dimensional PCA model. All p values represent the results of repeated measures t tests.

within-subjects factors are shown in Table S2, and the distributions of reconstruction error for different model types are shown in Figure 3A.

The type of model used, i.e. principal-component analyses or sequential autoencoder, has a significant effect on reconstruction accuracy. The sequential autoencoder produces significantly more accurate reconstructions than PCA (Table S3). The source of the data used for model training, i.e. individualized models using subjects' own data or transferred models trained using only data from other participants, also had a significant effect on reconstruction performance. Individualized models performed significantly better than transferred models (Table S4).

The sequential autoencoder significantly outperforms PCA for both individualized and transferred models (Table S5). Using transferred rather than individualized models results in a deterioration in performance for both PCA and sequential autoencoder models. However, sequential autoencoders trained on other participants' data showed equivalent performance to individualized PCA models, indicating greater generalizability of the features learned by sequential autoencoders.

The improved performance of the sequential autoencoder can be attributed to the use of temporal information. As shown in Figure 3B, an equivalent autoencoder without a recurrent network component demonstrates reconstruction performance at the same level as PCA, whereas the sequential autoencoder outperforms both (Table S6). This suggests that the improved performance is due to the ability to capture temporal dynamics in the movement data.

The accuracy of the reconstruction obtained increases with the number of latent features used for reconstruction. This is shown in Figure 3C. For both PCA and SAE, as decreasing numbers of features are used to describe the data, the ability to accurately reconstruct accurate hand movements also decreases. However, sequential autoencoders are able to perform more accurate reconstructions with lower-dimensional representations than standard PCA models. Superior performance to a PCA model with a 9-dimensional representation is achieved using only 5 features with a sequential autoencoder.

Nonlinear models that account for temporal dynamics outperform simple linear models in learning useful features. Figure S1 shows reconstructions produced by linear and nonlinear models for an example trial

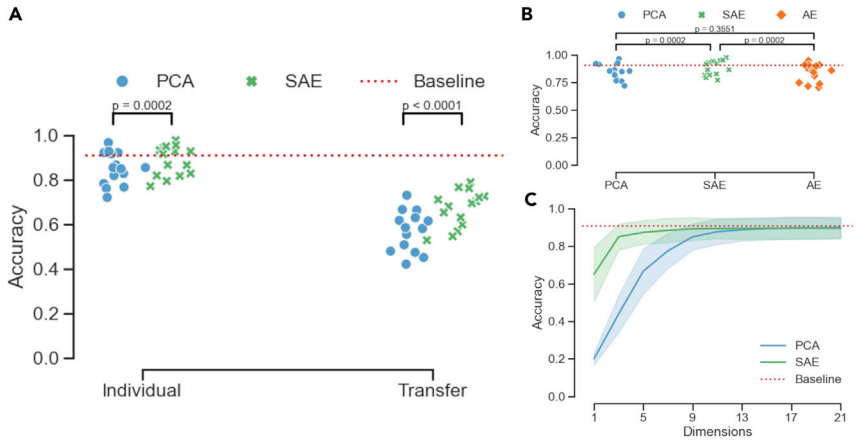


Figure 4. Reconstructed movements maintain functional distinctions

(A) Classification accuracy for data reconstructed using principal component analysis (PCA) and sequential autoencoder (SAE) models; the SAE is significantly better at preserving features that distinguish movements classes. This difference is more marked when using transferred models. Transferred models result in a deterioration in performance for both model types, although this effect is less marked for SAE models.

(B) MSE loss for PCA, SAE, and a standard autoencoder (AE) without temporal information. Removal of temporal dynamics in the AE model results in equivalent performance to PCA, whereas the SAE outperforms both. The improved performance of the SAE model is due to the use of temporal information.

(C) Classification accuracy of data reconstructed using models with a range of dimensionalities; the SAE performs significantly better at maintaining distinguishing features, achieving very high accuracies even with very few features. A 3-dimensional SAE model achieves superior performance to a 9-dimensional PCA model. All p values represent the results of repeated measures t tests.

using both individualized and transferred data. The sequential autoencoder produces more accurate reconstructions. This is also true for models that are trained on other participants' data; the sequential autoencoder is better able to learn latent features that generalize to new participants and accurately reconstruct their movements, whereas PCA fails to produce meaningful reconstructions using transferred models.

Temporal dynamics improve movement classifications

Kinematic data reconstructed from latent features retain the functional distinctions required to classify movement types. The effects of model type and data source on classification accuracy of reconstructed data are shown in Table S7. The distributions of classification accuracy for PCA and sequential autoencoders using individualized and transferred models are shown in Figure 4A.

Data reconstructed using sequential autoencoders can be classified into functional movement types significantly more accurately than data reconstructed using PCA, indicating greater recovery of functionally important movement features (Table S8). As in the case of movement reconstructions, individualized models provide superior classification performance than models trained on other participants' data (Table S9). The interaction between model type and data source also has a significant effect on classifier performance, indicating that the performance drop when using transferred models is not equal for PCA and sequential autoencoders (Table S10). Sequential autoencoders significantly outperform PCA for both individualized and transferred models.

The improved performance of the sequential autoencoders can be attributed to the use of temporal information. As shown in Figure 4B, an autoencoder with the recurrent network removed produced equivalent performance to PCA, whereas the sequential autoencoder outperforms both (Table S11). This indicates that the performance gains are due to the use of temporal dynamics in the movement data.

Functional distinctions are lost as the number of features used to reconstruct movements decreases. Figure 4C shows classification accuracy for reconstructed data for models with varying numbers of features. This demonstrates that below a certain point, functionally important features that differentiate movement types start to be lost. Notably, the sequential autoencoder achieves higher accuracies with

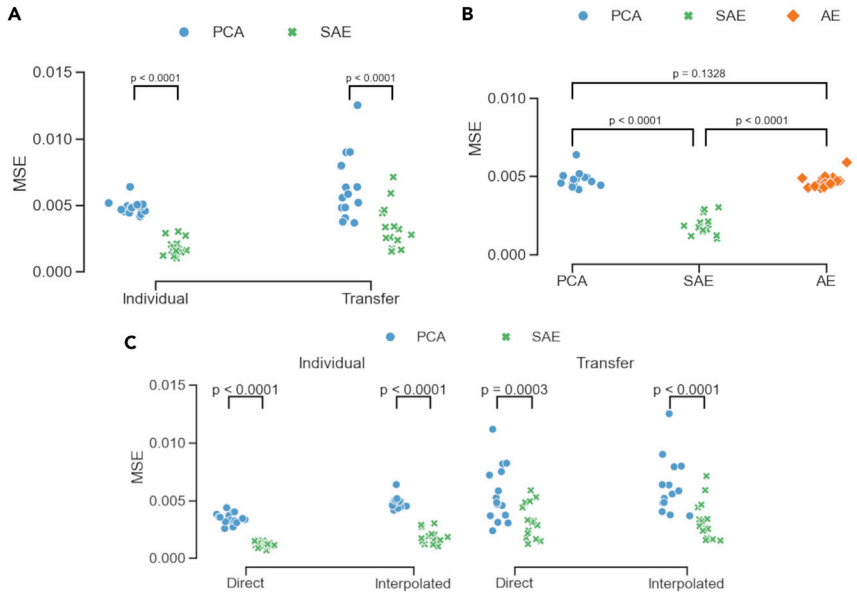


Figure 5. New movements can be interpolated using latent features

(A) Mean squared error (MSE) reconstruction loss for interpolated movements using principal component analysis (PCA) and sequential autoencoder (SAE) models using individualized and transferred data sources. The SAE model performs significantly better at interpolating unseen movements with both data sources. Using transferred models results in a significant loss of interpolation performance. Transferred SAE models produce interpolations of superior accuracy to individualized PCA models.

(B) MSE loss for PCA, SAE, and a standard autoencoder (AE) without temporal information. Removal of temporal information in the AE model produces equivalent performance to PCA, whereas the SAE model outperforms both. The improved interpolation performance of the SAE model is due to the use of temporal information.

(C) Reconstruction loss for directly reconstructed and interpolated movements. Using individualized models, sequential autoencoders (SAE) produce more accurate direct and interpolated reconstructions than principal component analysis (PCA). Interpolation results in a decrease in reconstruction accuracy for both model types. Interpolations performed using SAE models were more accurate than direct reconstructions using PCA models. Using transferred data, SAE produce more accurate direct and interpolated movements. Interpolation results in a decrease in performance for PCA, whereas SAE models reconstruct unseen movements as accurately as direct reconstructions. Interpolations performed with SAE models are more accurate than direct reconstructions using PCA models. All p values represent the results of repeated measures t tests.

lower-dimensional representations than PCA, with only 3 features required to achieve superior performance to a PCA model with a 9-dimensional representation. This indicates that these nonlinear models are better able to capture features that distinguish movement classes.

A range of classifiers trained on full sets of kinematic data are able to classify data reconstructed from latent features with high accuracy, as shown in Figure S2. The highest classification accuracy was 85.0% for PCA and 88.6% for sequential autoencoders. The baseline accuracy achieved on the full kinematic data without reconstruction was 91.9%, representing only a 3.3 percentage point reduction in performance using data reconstructed using a sequential autoencoder. The sequential autoencoder produces superior reconstruction accuracies using all classifiers tested.

Dynamic features allow interpolation of new movements

New, previously unseen movements can be reconstructed by interpolation using features learned on other movements. By holding out all instances of a given movement from the training set, training on all other data, and then testing reconstruction on the held-out movement, the ability of latent features to generalize to new movements can be assessed. This allows us to assess whether the features learned on a subset of movements are able to accurately represent and reconstruct previously unseen movements, indicating that the latent features capture the general structure of hand movements.

The effects of model type and data source on interpolation error are shown in Table S12, and the distributions of interpolation error for different model types are shown in Figure 5A.

The type of model used has a significant effect on interpolation accuracy. The sequential autoencoder produces significantly more accurate interpolations of unseen movements than PCA (Table S13). The source of the data used for model training also had a significant effect on interpolation performance. Individualized models produced superior interpolations to transferred models (Table S14).

The sequential autoencoder produces better interpolations of unseen movements than PCA for both individualized and transferred models (Table S15). Using transferred models results in a significant drop in interpolation performance for both PCA and SAE models. Transferred sequential autoencoders demonstrated superior performance to individualized PCA models, highlighting the superior ability of features learned using sequential autoencoders to transfer to new settings outside the training context.

The improved performance of sequential autoencoders can be attributed to the use of temporal information. Figure 5B shows a comparison of interpolation performance for sequential autoencoders, PCA, and an equivalent autoencoder without a recurrent stage. The standard autoencoder and PCA perform similarly, whereas the sequential autoencoder outperforms both (Table S16). This indicates that the performance gain with the sequential autoencoder is due to the use of temporal dynamics in the movement data.

Figure 5C shows a direct comparison of directly reconstructed (i.e. trained on data including that movement) and interpolated (i.e. trained on data that do not include that movement) movements for each model type and data source. Using individualized models, interpolated reconstructions are less accurate than direct reconstructions for both PCA and sequential autoencoders (Table S17). Using transferred models, interpolation results in a significant drop in performance compared with direct reconstruction when using PCA, but there is no reduction in performance for sequential autoencoders, indicating equivalent performance for previously unseen movements when using models trained on other participants' data. This further highlights the ability of sequential autoencoders to learn features that generalize across participants and movements. Further, interpolated movements using a sequential autoencoder were more accurate than directly reconstructed movements using PCA for individualized and transferred models, indicating that sequential autoencoders can provide better reconstructions of previously unseen movements than PCA can of movements in the training data.

Interpolation of previously unseen movements has an effect on reconstruction performance. Figure S3 shows examples of direct and interpolated reconstructions using PCA and sequential autoencoders using individualized and transferred models. The sequential autoencoder produces more accurate interpolations in all cases and provides minimal loss of performance when interpolating unseen movements in comparison to reconstructing movements from the training data. Notably, PCA demonstrates a significant drop in performance when using transferred models, and interpolations of unseen movements using transferred models contain little useful information. Conversely, sequential autoencoders produce reasonable reconstructions when using transferred models, including when interpolating unseen movements using a model trained on other participants' data.

Generalizable features can be learned with limited training data

By measuring interpolation error for each movement, we can rank movements based on the extent to which they can be interpolated based on features derived from other movements. This gives us a measure of how informative each movement is in learning latent features, i.e. how much unique information is contained in that movement. The list of movements ranked by interpolation loss (with the top having the highest error, i.e. the most difficult to interpolate) is shown in Table S18. The rankings of movements across model types is largely consistent, with much of the top and bottom five shared across models. This provides a useful resource for identifying the most informative movements to include in a training dataset for maximizing reconstruction accuracy.

Useful latent features can be learned from limited sets of training tasks. By ranking the movements used according to their interpolation error, models were trained on a range of subsets of the training tasks. In each case, the top *n* tasks were selected from the training data and models were trained on this subset and evaluated on the full test set with all tasks included. This was performed for models using only the top task to models using all tasks. The number of training movements required to achieve useful reconstructions was then assessed.

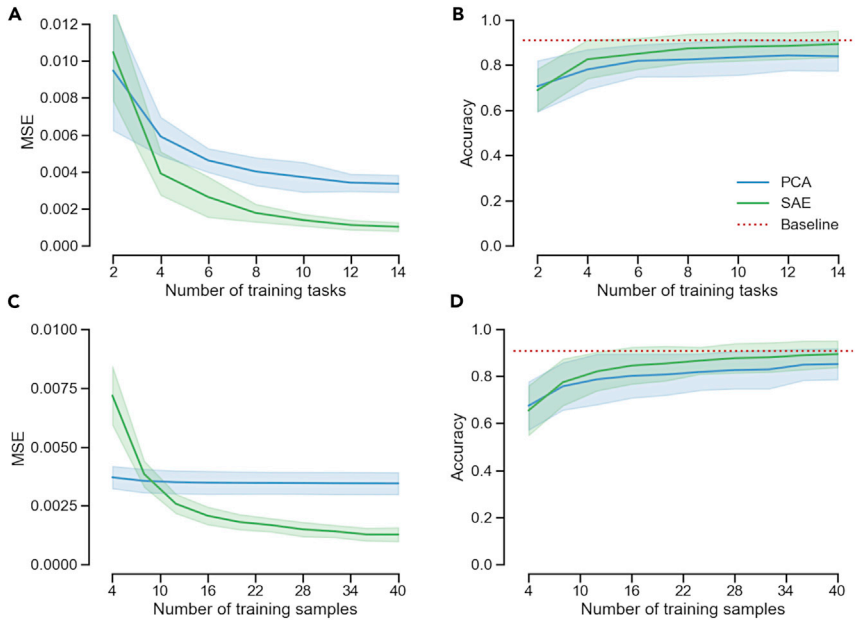


Figure 6. Movement features can be characterized using a limited set of training data

(A) Mean squared error (MSE) reconstruction loss for principal component analysis (PCA) and sequential autoencoder (SAE) models on all tasks when trained on a subset of tasks. Reconstruction accuracy increases as more task types are included in training. The SAE model achieves high reconstruction accuracy on all tasks when only trained on a small number of tasks. An SAE model trained on 4 tasks achieves equivalent performance to PCA models trained on 10 tasks, whereas an SAE model trained on greater than 6 tasks achieves superior performance across all tasks to PCA models trained on all tasks.

(B) Classification accuracy of reconstructed data using models trained on a subset of tasks. Classification accuracy improves as the number of training tasks is increased. High classification accuracy, including for unseen tasks, is achieved even when only trained on a small number of tasks.

(C) MSE for PCA and SAE models when trained on a limited number of samples of each movement type. PCA performance is relatively static as training examples are increased beyond the first small number of samples. SAE performance increases as more samples of each movement are included, achieving superior performance to PCA with 12 examples of each movement.

(D) Classification accuracy of reconstruction data using models trained on limited numbers of samples of each movement type. Classification accuracy improves as the number of training samples is increased. High classification accuracies are achieved even with low numbers of training samples.

The reconstruction error produced for a range of sets of training tasks is shown in Figure 6A. Sequential autoencoder models produce more accurate reconstructions with more limited training sets than PCA, indicating a superior ability to interpolate movements outside the training set. To achieve the same reconstruction performance on all tasks as a PCA model trained using 10 tasks, only 4 training tasks were required to train the sequential autoencoder. Sequential autoencoders trained with greater than 6 training tasks achieved superior performance across all tasks than PCA models trained on all tasks.

The classification accuracy for a range of sets of training tasks is shown in Figure 6B. The sequential autoencoder similarly produces superior performance with a more limited set of training tasks. These results demonstrate that it is possible to derive useful latent features with only a very limited set of training tasks and to use these features to achieve good performance on a wide range of movement tasks through interpolation.

The ability of models to learn useful features from limited numbers of training samples was assessed by restricting the number of examples of each movement type available in the training data. The reconstruction and classification accuracies for models trained on limited numbers of movement samples are shown in Figures 6C and 6D. PCA performance is relatively static as the number of movement samples is increased. SAE performance, in terms of both reconstruction and classification, increases rapidly with the amount of available training data. Useful features are learned with even small amounts of training data, achieving superior performance to PCA models trained on 40 samples of each movement with only 12 samples of each

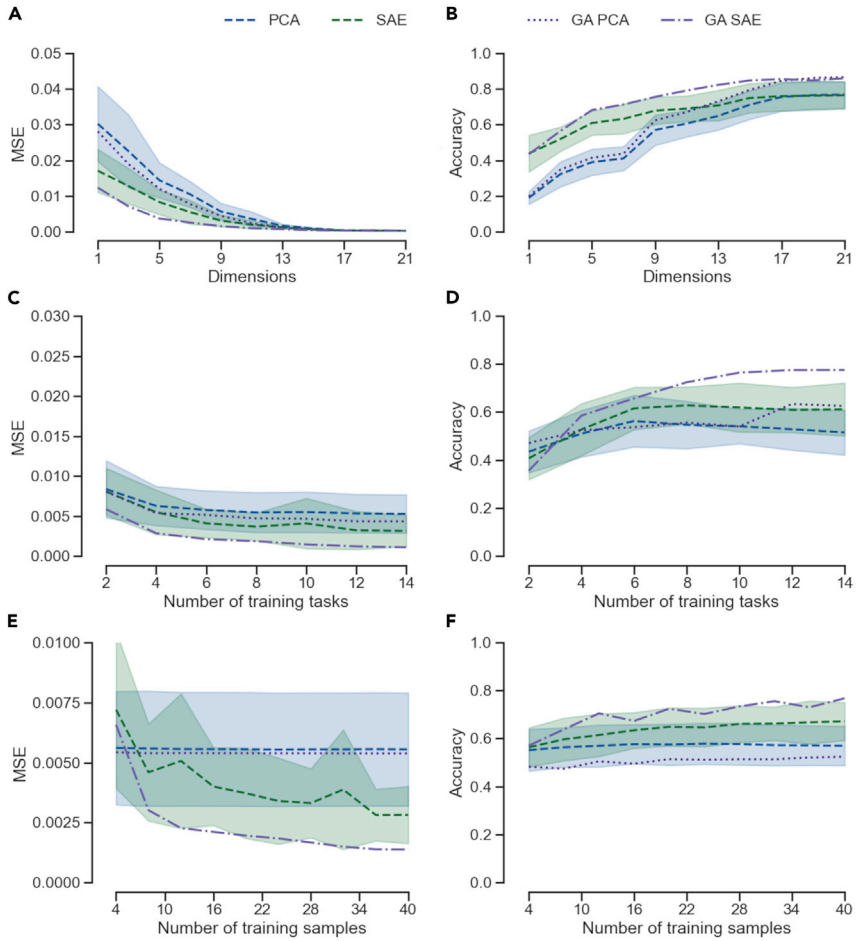


Figure 7. Movement features are generalizable and transferable between individuals

(A) Mean squared error (MSE) reconstruction loss for principal component analysis (PCA) and sequential autoencoder (SAE) models with varying numbers of features using transferred data, with models trained on all data (grand-average; GA) for comparison. Accurate reconstructions are achieved on unseen participants' data even with relatively low-dimensional representations.

(B) Classification accuracy of data reconstructed using models with varying numbers of features. High classification accuracies are achieved even with low-dimensional representations using other participants' data.

(C) Reconstruction loss using models trained on a subset of training tasks using other participants' data. Accurate reconstructions across all tasks on unseen participants' data are achieved even when only trained on a small set of tasks.

(D) Classification accuracy of data reconstructed using models trained on a subset of tasks tested on unseen participants' data; high accuracies are achieved across all tasks even when only trained on a subset of tasks.

(E) MSE for PCA and SAE models trained on limited numbers of samples of each movement type. PCA benefits little from additional data, whereas SAE model performance increases with additional movement samples. Superior reconstruction performance to PCA models trained on 40 samples are achieved with only 8 training samples using a SAE model.

(F) Classification accuracy of data reconstructed using models trained on limited sets of training samples. High accuracies are achieved even with only small numbers of training samples. SAE models consistently outperform PCA models.

movement. These results demonstrate that the number of training samples available significantly improves SAE performance but that useful features can be learned with even very few movement examples.

Dynamic features are conserved across individuals

Movement features generalize to populations and can be learned on populations and applied to individuals. The ability of learned features to generalize to a population was assessed using a grand-average analysis. All participants' data were pooled and analyzed together. The performance of features learned on a population rather than an individual was qualitatively similar to individual models. The reconstruction error and classification accuracy produced over a range of feature sizes are shown in Figures 7A and 7B, respectively. The ability

to identify generalizable features on a population using a limited training set was also qualitatively similar to individual models. The reconstruction error and classification accuracy produced by features learned from training sets of varying size are shown in Figures 7C and 7D respectively. Sequential autoencoder models demonstrate superior performance, with more accurate movement reconstructions with both smaller feature sets and more limited training data. Similarly, Figures 7E and 7F show reconstruction error and classification performance produced by features learned using limited numbers of training samples of each movement. Sequential autoencoders rapidly learn useful features, with only 8 samples of each movement required to outperform a PCA model trained on 40 samples of each movement. Sequential autoencoder rapidly learn useful representations with very little data, significantly outperforming PCA.

The ability of population-based models to provide accurate results for a new individual was assessed by training models on the data of all other participants and testing on the held-out participant; this was repeated for all participants in order to evaluate the ability of generalized models to accurately reconstruct the movements of an individual not included in the training set. The reconstruction error produced by population-based models applied to individuals is shown in Figure 7A. Sequential autoencoder models provide superior performance with fewer features, with only a 7-dimensional representation required to achieve superior performance to a PCA model with a 9-dimensional representation. Similarly, the classification accuracy produced over a range of feature sizes is shown in Figure 7B. Sequential autoencoders demonstrate superior performance, with only 3 features required to achieve equivalent accuracy to a PCA model with a 9-dimensional representation.

Figure 7C demonstrates the reconstruction error for models trained using subsets of the training data for population-based models. This demonstrates that population-based features can still be learned using only a small number of training tasks and can still generalize well to unseen movements in an unseen participant. Sequential autoencoders outperform PCA even with very limited sets of training data, with only 4 training tasks required to achieve superior performance on the whole dataset to PCA models trained using all tasks. The classification accuracy achieved over a range of training task sets is shown in Figure 7D. This shows a similar pattern, with sequential autoencoders producing consistently superior performance with limited training sets. This further demonstrates that practically useful features can be derived from a limited training set on a population and can generalize well to a new individual without the need for individualized training.

The effect of limited numbers of training samples on model performance is shown in Figures 7E and 7F. This demonstrates that useful population-based features can be learned using very few training examples. Sequential autoencoders rapidly learn useful features, outperforming a PCA model trained on 40 examples of each movement in terms of both reconstruction error and classification accuracy with only 8 examples of each. This demonstrates that useful features that generalize to a population can be learned using only a very limited number of training examples using a sequential autoencoder.

DISCUSSION

The extent of hand synergies and the intrinsic dimensionality of hand movements remains a topic of debate. Recent evidence on the application of nonlinear dimensionality reduction techniques has shown that the addition of complex nonlinearities does not by itself capture the underlying structure of hand movements more accurately than linear methods.[31,32] This has been interpreted as suggesting that hand kinematics may not be as low-dimensional as previously thought. However, our results demonstrate that by incorporating consideration of the temporal dynamics of hand movements in line with approaches used on the neural representations of hand movements, we are able to more accurately capture the structure of movements, with most task-relevant information being contained in a very small number of components. These dynamic features allowed greater reconstruction accuracy and generalization with smaller feature sets than linear models. This suggests that hand movements are strongly low-dimensional and that the latent features that comprise the space of hand movements have strong temporal dynamics. This strong spatiotemporal structure is further underlined by the need for only a small number of movement examples to learn a useful set of dynamic features that can accurately reconstruct all tested movements.

Although PCA can identify sufficient structure to provide accurate reconstructions with a low-dimensional representation within an individual, these features have a limited ability to interpolate unseen movements, they do not transfer well to new individuals, and they fail to produce useful interpolations of unseen movements in new individuals. Features incorporating temporal dynamics, by contrast, accurately

interpolate new movements, allow use of learned features in new individuals with little impact on reconstruction accuracy, and can interpolate unseen movements in new individuals with no further drop in performance. The movements are quickly learned on small sets of training data and quickly converge on a conserved set of behaviors across individuals. This suggests that the learned features are strongly conserved across movements and across individuals, possibly reflecting the actual movement primitives encoded by the CNS.

This latent dynamical structure is in line with neurobiological evidence for hand synergies. Individual cells in the primary motor cortex do not operate individual muscles.[21–24] Rather, individual neurons produce complex patterns of muscle activations mediated by circuits of spinal interneurons.[25] Further, the hand is controlled by direct corticomotoneuronal pathways,[5,26–30] which directly produce specific patterns of muscle co-activation. These synergistic activations of multiple muscles represent candidates for the "synergies" exploited by our sequential autoencoder model.

Synergies with temporal dynamics such as those identified here are in keeping with theories describing primary motor cortex as a dynamical system or pattern generator,[37] where current state predicts future state. These dynamical features can then define a time-dependent manifold,[14] allowing low-dimensional representation of complex movements using synergies with temporal features.[10,38] Interestingly, motor cortex dynamics appear to differ for reaching and grasping movements, with variable use of static and spatiotemporal synergies depending on movement type.[39,40] Interestingly, the ability to use generalized motor synergies to generate new movements has recently been demonstrated in a computational model of reach-and-grasp movements[41]; our results demonstrate this phenomenon in kinematic data of functional hand movements.

Limitations of the study

Although we have shown that it is possible to achieve accurate reconstructions of functional hand movements using a very small number of dynamical features, this may not be sufficient for optimal control. In particular, very fine motor tasks involving individual finger movement, required for example to play a piano, were not included in the tasks tested. It may be that such fine movements require additional levels of control, with more complex representations. Further, our functional movements did not include any interactions with objects. The hand subtly alters its shape in order to conform to objects it is interacting with,[36,42] and these involuntary adaptations may not be captured with this kind of approach. Indeed, it was shown previously that the information content of low-variance principal components was related to these kinds of subtle movements.[31] Further assessment of the ability of dynamical features to capture movement structure during object interactions would be required in order to properly investigate whether these small adaptive movements increase the intrinsic dimensionality of hand movements.

Conclusion

The ability to capture the underlying structure of hand kinematics and represent this efficiently has important practical applications. By understanding the structure of functional hand movements, we are better equipped to develop technologies for the restoration of lost function. Using a generative approach, the sequential autoencoder model could be used to produce realistic, functional hand movements by interpolation on the manifold of dynamic hand synergies, potentially providing a useful control system for restorative neuroprostheses.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participant recruitment
- METHOD DETAILS
  - Experiment design
  - Data collection

- Task
- Data preprocessing
- Linear dimensionality reduction
- Non-linear dimensionality reduction
- Ablation study
- Reconstruction accuracy
- Classification accuracy
- Movement interpolation
- Assessing generalization

- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis
  - Software

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2022.105428.

ACKNOWLEDGMENTS

This work was supported by the National Institute for Health Research (NIHR) through the NIHR Oxford Biomedical Research Centre and the UK Engineering and Physical Sciences Research Council (EPSRC) via the University of Oxford Clarendon Fund. The views expressed are those of the authors and not necessarily those of the funding bodies.

AUTHOR CONTRIBUTIONS

C.K. and J.F. conceived and designed the analysis. C.K. collected the data and performed the analysis. C.K. and J.F. wrote the paper. J.F. supervised the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: May 2, 2022
Revised: September 1, 2022
Accepted: October 19, 2022
Published: November 18, 2022

REFERENCES

1. Sobinov, A.R., and Bensmaia, S.J. (2021). The neural mechanisms of manual dexterity. Nat. Rev. Neurosci. *22*, 741–757. https://doi.org/10.1038/s41583-021-00528-7.

2. Lee, J.C., and Healy, J.C. (2005). Normal sonographic anatomy of the wrist and hand. Radiographics *25*, 1577–1590. https://doi.org/10.1148/RG.256055028.

3. Johnson, K.O. (2001). The roles and functions of cutaneous mechanoreceptors. Curr. Opin. Neurobiol. *11*, 455–461. https://doi.org/10.1016/S0959-4388(00)00234-8.

4. Sur, M., Merzenich, M.M., and Kaas, J.H. (1980). Magnification, receptive-field area, and "hypercolumn" size in areas 3b and 1 of somatosensory cortex in owl monkeys. J. Neurophysiol. *44*, 295–311. https://doi.org/10.1152/JN.1980.44.2.295.

5. Rathelot, J.A., and Strick, P.L. (2009). Subdivisions of primary motor cortex based on cortico-motoneuronal cells. Proc. Natl. Acad. Sci. USA *106*, 918–923. https://doi.org/10.1073/PNAS.0808362106.

6. Kutch, J.J., and Valero-Cuevas, F.J. (2012). Challenges and new approaches to proving the existence of muscle synergies of neural origin. PLoS Comput. Biol. *8*, e1002434. https://doi.org/10.1371/JOURNAL.PCBI.1002434.

7. Santello, M., Bianchi, M., Gabiccini, M., Ricciardi, E., Salvietti, G., Prattichizzo, D., Ernst, M., Moscatelli, A., Jörntell, H., Kappers, A.M.L., et al. (2016). Hand synergies: integration of robotics and neuroscience for understanding the control of biological and artificial hands. Phys. Life Rev. *17*, 1–23. https://doi.org/10.1016/J.PLREV.2016.02.001.

8. Todorov, E., and Ghahramani, Z. (2004). Analysis of the synergies underlying complex hand manipulation. In Conference Proceedings IEEE Engineering in Medicine and Biology Society, *2004* (IEEE), pp. 4637–4640. https://doi.org/10.1109/IEMBS.2004.1404285.

9. D'Avella, A., Portone, A., Fernandez, L., and Lacquaniti, F. (2006). Control of fast-reaching movements by muscle synergy combinations. J. Neurosci. *26*, 7791–7810. https://doi.org/10.1523/JNEUROSCI.0830-06.2006.

10. Overduin, S.A., d'Avella, A., Roh, J., Carmena, J.M., and Bizzi, E. (2015). Representation of muscle synergies in the primate brain. J. Neurosci. *35*, 12615–12624. https://doi.org/10.1523/JNEUROSCI.4302-14.2015.

11. Flash, T., and Hochner, B. (2005). Motor primitives in vertebrates and invertebrates. Curr. Opin. Neurobiol. *15*, 660–666. https://doi.org/10.1016/J.CONB.2005.10.011.

12. Tresch, M.C., and Jarc, A. (2009). The case for and against muscle synergies. Curr. Opin. Neurobiol. *19*, 601–607. https://doi.org/10.1016/J.CONB.2009.09.002.

13. Ingram, J.N., Körding, K.P., Howard, I.S., and Wolpert, D.M. (2008). The statistics of natural hand movements. Exp. Brain Res. *188*, 223–236. https://doi.org/10.1007/S00221-008-1355-3.

14. Gallego, J.A., Perich, M.G., Miller, L.E., and Solla, S.A. (2017). Neural manifolds for the control of movement. Neuron *94*, 978–984. https://doi.org/10.1016/J.NEURON.2017.05.025.

15. Berniker, M., Jarc, A., Bizzi, E., and Tresch, M.C. (2009). Simplified and effective motor control based on muscle synergies to exploit musculoskeletal dynamics. Proc. Natl. Acad. Sci. USA *106*, 7601–7606. https://doi.org/10.1073/PNAS.0901512106.

16. d'Avella, A., and Lacquaniti, F. (2013). Control of reaching movements by muscle synergy combinations. Front. Comput. Neurosci. 42. https://doi.org/10.3389/FNCOM.2013.00042/BIBTEX.

17. Ringnér, M. (2008). What is principal component analysis? Nat. Biotechnol. *26*, 303–304. https://doi.org/10.1038/nbt0308-303.

18. Tresch, M.C., Cheung, V.C.K., and D'Avella, A. (2006). Matrix factorization algorithms for the identification of muscle synergies: evaluation on simulated and experimental data sets. J. Neurophysiol. *95*, 2199–2212. https://doi.org/10.1152/JN.00222.2005.

19. Furuya, S., Flanders, M., and Soechting, J.F. (2011). Hand kinematics of piano playing. J. Neurophysiol. *106*, 2849–2864. https://doi.org/10.1152/JN.00378.2011.

20. Santello, M., Flanders, M., and Soechting, J.F. (1998). Postural hand synergies for tool use. J. Neurosci. *18*, 10105–10115. https://doi.org/10.1523/JNEUROSCI.18-23-10105.1998.

21. Overduin, S.A., d'Avella, A., Carmena, J.M., and Bizzi, E. (2012). Microstimulation activates a handful of muscle synergies. Neuron *76*, 1071–1077. https://doi.org/10.1016/J.NEURON.2012.10.018.

22. Schieber, M.H. (2001). Constraints on somatotopic organization in the primary motor cortex. J. Neurophysiol. *86*, 2125–2143. https://doi.org/10.1152/JN.2001.86.5.2125.

23. Capaday, C., Ethier, C., Van Vreeswijk, C., and Darling, W.G. (2013). On the functional organization and operational principles of the motor cortex. Front. Neural Circuits *7*, 66. https://doi.org/10.3389/FNCIR.2013.00066/BIBTEX.

24. Goodman, J.M., Tabot, G.A., Lee, A.S., Suresh, A.K., Rajan, A.T., Hatsopoulos, N.G., and Bensmaia, S. (2019). Postural representations of the hand in the primate sensorimotor cortex. Neuron *104*, 1000–1009.e7. https://doi.org/10.1016/j.neuron.2019.09.004.

25. Takei, T., Confais, J., Tomatsu, S., Oya, T., and Seki, K. (2017). Neural basis for hand muscle synergies in the primate spinal cord. Proc. Natl. Acad. Sci. USA *114*, 8643–8648. https://doi.org/10.1073/PNAS.1704328114/-/DCSUPPLEMENTAL.

26. Lemon, R.N. (2008). Descending pathways in motor control. Annu. Rev. Neurosci. *31*, 195–218. https://doi.org/10.1146/ANNUREV.NEURO.31.060407.125547.

27. McKiernan, B.J., Marcario, J.K., Karrer, J.H., and Cheney, P.D. (1998). Corticomotoneuronal postspike effects in shoulder, elbow, wrist, digit, and intrinsic hand muscles during a reach and prehension task. J. Neurophysiol. *80*, 1961–1980. https://doi.org/10.1152/JN.1998.80.4.1961.

28. Griffin, D.M., Hoffman, D.S., and Strick, P.L. (2015). Corticomotoneuronal cells are "functionally tuned. Science *350*, 667–670. https://doi.org/10.1126/SCIENCE.AAA8035.

29. Muir, R.B., and Lemon, R.N. (1983). Corticospinal neurons with a special role in precision grip. Brain Res. *261*, 312–316. https://doi.org/10.1016/0006-8993(83)90635-2.

30. Lemon, R. (2019). Recent advances in our understanding of the primate corticospinal system. F1000Res. *8*. https://doi.org/10.12688/F1000RESEARCH.17445.1/.

31. Yan, Y., Goodman, J.M., Moore, D.D., Solla, S.A., and Bensmaia, S.J. (2020). Unexpected complexity of everyday manual behaviors. Nat. Commun. *11*, 3564. https://doi.org/10.1038/s41467-020-17404-0.

32. Patel, V., and Burns, M. (2015). Linear and nonlinear kinematic synergies in the grasping hand. J. Bioeng Biomed. Sci. *05*. https://doi.org/10.4172/2155-9538.1000163.

33. Santello, M., Baud-Bovy, G., and Jörntell, H. (2013). Neural bases of hand synergies. Front. Comput. Neurosci. *7*, 23. https://doi.org/10.3389/FNCOM.2013.00023.

34. Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural Comput. *9*, 1735–1780. https://doi.org/10.1162/NECO.1997.9.8.1735.

35. Pandarinath, C., O'Shea, D.J., Collins, J., Jozefowicz, R., Stavisky, S.D., Kao, J.C., Trautmann, E.M., Kaufman, M.T., Ryu, S.I., Hochberg, L.R., et al. (2018). Inferring single-trial neural population dynamics using sequentialauto-encoders. Nat. Methods *15*, 805–815. https://doi.org/10.1038/S41592-018-0109-9.

36. Thakur, P.H., Bastian, A.J., and Hsiao, S.S. (2008). Multidigit movement synergies of the human hand in an unconstrained haptic exploration task. J. Neurosci. *28*, 1271–1281. https://doi.org/10.1523/JNEUROSCI.4512-07.2008.

37. Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., Shenoy, K.V., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56. https://doi.org/10.1038/nature11129.

38. D'Avella, A., Saltiel, P., and Bizzi, E. (2003). Combinations of muscle synergies in the construction of a natural motor behavior. Nat. Neurosci. *6*, 300–308. https://doi.org/10.1038/NN1010.

39. Suresh, A.K., Goodman, J.M., Okorokova, E.V., Kaufman, M., Hatsopoulos, N.G., and Bensmaia, S.J. (2020). Neural population dynamics in motor cortex are different for reach and grasp. Elife *9*, e58848. https://doi.org/10.7554/ELIFE.58848.

40. Rouse, A.G., and Schieber, M.H. (2018). Condition-dependent neural dimensions progressively shift during reach to grasp. Cell Rep. *25*, 3158–3168.e3. https://doi.org/10.1016/j.celrep.2018.11.057.

41. Kutsuzawa, K., and Hayashibe, M. (2022). Motor synergy generalization framework for new targets in multi-planar and multi-directional reaching task. R. Soc. Open Sci. *9*, 211721. https://doi.org/10.1098/RSOS.211721.

42. Santello, M., Flanders, M., and Soechting, J.F. (2002). Patterns of hand motion during grasping and the influence of sensory guidance. J. Neurosci. *22*, 1426–1435. https://doi.org/10.1523/JNEUROSCI.22-04-01426.2002.

43. Light, C.M., Chappell, P.H., and Kyberd, P.J. (2002). Establishing a standardized clinical assessment tool of pathologic and prosthetic hand function: normative data, reliability, and validity. Arch. Phys. Med. Rehabil. *83*, 776–783. https://doi.org/10.1053/APMR.2002.32737.

44. Kingma, D.P., and Ba, J.L. (2014). Adam: a method for stochastic optimization. Preprint at arXiv. 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. https://doi.org/10.48550/arXiv.1412.6980.

45. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2.

46. Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

47. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury Google, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). PyTorch: An Imperative Style (High-Performance Deep Learning Library).

48. Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. PROC. 9th PYTHON Sci. CONF.

49. Hunter, J.D. (2007). Matplotlib: a 2D graphics environment. Comput. Sci. Eng. *9*, 90–95. https://doi.org/10.1109/MCSE.2007.55.

50. Waskom, M. (2021). seaborn: statistical data visualization. J. Open Source Softw. *6*, 3021. https://doi.org/10.21105/JOSS.03021.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python 3.7	Python Software Foundation	http://www.python.org/
Numpy	NumPy steering council	https://numpy.org/
scikit-learn	Scikit-Learn team	https://scikit-learn.org/stable/
PyTorch	PyTorch Foundation	https://pytorch.org/
Sequential autoencoders for hand kinematics	Oxford Neural Interfacing	https://github.com/Oxford-Neural-Interfacing/handkinematics; https://doi.org/10.5281/zenodo.7213049

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, James FitzGerald (james.fitzgerald@nds.ox.ac.uk).

Materials availability

This study did not generate any new unique reagents.

Data and code availability

All data reported in this paper will be shared by the lead contact upon request.

All analysis code is available at https://github.com/Oxford-Neural-Interfacing/handkinematics. Code has also been deposited at Zenodo with a DOI listed in the key resources table. All code is publicly available at the time of publication.

Any additional information required to reanalyse the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participant recruitment

Healthy, right-handed volunteers with no musculoskeletal or neurological pathologies were recruited. 15 participants (9 male, 6 female) were recruited. Participants consented to taking part in the study and to retention of study data. All procedures were approved by local ethics committees and were carried out in line with local ethical guidelines for studies involving human participants.

METHOD DETAILS

Experiment design

Detailed kinematic data was recorded during functional hand movements in healthy volunteers. A non-linear dimensionality reduction method was developed to learn features with temporal dynamics. In order to evaluate whether these dynamic features captured the structure of hand kinematics, this model was compared to a linear dimensionality reduction approach that is widely used in the field, i.e. principal component analysis. The ability of standard linear and non-linear dynamic features to capture the variability in the data was evaluated and compared. Models were then compared based on their ability to accurate reconstruct the original kinematic data from a limited feature set, to produce movement reconstructions that allow accurate movement classification and to generalize to new, previously unseen movements. The ability of learned features to generalize to new participants was assessed by training models on other participants' data, assessing the ability of these approaches to learn features that are conserved across individuals, reflecting a shared underlying structure of hand movements rather than individual idiosyncrasies.

Data collection

Hand kinematic data was measured using a data glove (Bebop Sensors) with bend sensors at the base and tip of each finger, an inertial measurement unit, accelerometer and two-axis goniometers (Biometrics) on the wrist and thumb, providing 21 channels of kinematic data. All data were sampled at 1024Hz.

Task

Participants were seated in front of a computer monitor with all sensors in place. Images of hand movements were displayed on screen and participants were asked to perform the hand movement shown. 14 functional hand positions were used, comprising a combination of functionally important grips and hand positions. These comprised the grip types used as part of the Southampton Hand Assessment Procedure[43] and simple positional movements to allow representation of changes of hand position and gestures. Table S1 provides a list of all movements used. Figure 2A shows the images seen by participants to demonstrate movements.

Following a training protocol where participants were familiarized with the task and the hand movements, hand movement images were presented randomly to the participant. Each trial lasted two seconds, during which the participant performed the movement and returned to a neutral position with hand flat on the table. Participants performed a total of 900 trials.

Data preprocessing

All data were downsampled to 128Hz. All trials with less than two seconds of data, representing an error with trial acquisition, were removed from analysis.

Data were divided into training, validation and test sets. For individual models, 60% of the participants data was used for training, 20% for validation and 20% for testing. For grand-average models, all participants' data was combined; 60% of the total population data was used for training, 20% for validation and 20% for testing. For analysis of the transfer of population models to individuals, all other participants' data was combined; 80% of this data was used for training and 20% for validation, while the participant's own data was used for testing. This was repeated for each participant.

All features in the training data were rescaled to the range 0–1. The same transformation (fit to the training data) was then applied to the test and validation data.

Linear dimensionality reduction

Movement features were identified using linear dimensionality reduction by applying probabilistic principal component analysis[17] to the high-dimensional kinematic data of the training set. This transforms the data into a new co-ordinate system using a singular value decomposition of the data in order to identify the features that explain most of the variability in the data. This is limited to features that are a linear combination of the recorded data and that are orthogonal to all other features, i.e. it simply describes a rotation and scaling to a new orthonormal basis. The derived features were ranked according to the variability in the original data they explained. All features ranked below the target number of features were discarded and the reduced set of features kept as a representation of kinematic data using a reduced feature set. Principal component analysis does not explicitly represent time, so data were transformed so that all timepoints were considered independent measurements of kinematics. As a consequence, the transformation is applied to the space of variability of hand kinematics without any representation of temporal dynamics rather than directly to the time-series data.

To allow reconstruction of detailed kinematic data from the reduced set of features, an inverse transformation was also computed. This allowed mapping from latent features to high-dimensional kinematics.

Non-linear dimensionality reduction

In order to better exploit any nonlinear dependencies between measurements of hand kinematics and to take advantage of the temporal structure of the kinematic time series data, a neural network architecture based on sequential autoencoders for learning the underlying dynamics of time series data was constructed.

A network consisting of an encoder, for mapping from hand kinematics to low-dimensional latent features, and a decoder, for mapping from low-dimensional latent features to hand kinematics, was constructed. By training the network to reconstruct its own input (i.e. the kinematic data), it learns the features in the bottleneck layer that allow maximal reconstruction of the original data. This allows the model to learn complex, non-linear dependencies between inputs that would not be captured by linear methods.

To account for the temporal structure of hand kinematic data, the encoder was made up of a long short-term memory recurrent neural network[34] in order to learn the dynamics of movements; this was connected to a fully connected network that mapped the output of the recurrent network to the low-dimensional latent features. The decoder was structured similarly, with a similar long short-term memory recurrent neural network and a fully connected network which maps the output of this network to detailed kinematic data. In this way, the non-linear and temporal structure of hand kinematics can be exploited to learn the optimal features required for reconstructing movements.

The encoder was therefore composed of a single LSTM layer with a hidden size of 64, followed by a fully connected layer with an output size of n, where n is the size of the latent space, i.e. the bottleneck layer. The decoder stage was made up of a single LSTM layer with a hidden size of 64, followed by a fully connected layer with an output size equivalent to the total number of dimensions in the original data. When the number of dimensions was changed, only the size of the bottleneck layer was altered; the model architecture otherwise remained unchanged.

The network was trained by updating the model weights to minimize the mean squared error between the original kinematic data and the kinematics reconstructed from latent features:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$$

where Y is the true kinematic data and $\widehat{Y}$ is the reconstructed kinematic data produced by the autoencoder.

Optimization was performed using the adaptive moment estimation (Adam) algorithm,[44] with a batch size of 32 for 250 training epochs.

Ablation study

In order to evaluate the contribution of the recurrent neural network to the performance of the sequential autoencoder, an equivalent model without the LSTM layer was constructed and trained using the same procedures. The performance of the two models could then be compared in order to assess the contribution of the use of temporal information to the results.

This LSTM ablation model comprised an encoder made up of a fully connected layer with an output size of 64, the same as the hidden size of the LSTM network, followed by another fully connected layer with an output size equivalent to the size of the bottleneck layer. The decoder was then made up of a fully connected layer with an output size of 64 and a second fully connected layer with an output size equivalent to the number of dimensions in the original data. In this way, the network was equivalent to the sequential autoencoder with the LSTM layer substituted for a fully connected layer.

Reconstruction accuracy

The ability to accurately reconstruct hand kinematics from sets of latent features was determined by applying a transformation to map the data from latent features to high-dimensional kinematics (i.e. the inverse transformation in the case of principal component analysis and the decoder in the case of the sequential autoencoder). Based on the existing literature in hand biomechanics, nine latent features was chosen as the standard to assess our models at, as it has been well demonstrated that hand kinematics can be mapped to a nine-dimensional feature space. The quality of the reconstruction was assessed by measuring the mean squared error between the actual data and the reconstructed data.

The impact of the number of features used on the reconstruction accuracy obtained was assessed by measuring the reconstruction accuracy achieved by models fit to a range of feature sets, from one up to the 21 (i.e. the total dimensionality of the input data).

Classification accuracy

In order to measure whether functionally important features that distinguish between movement types were preserved by the features used, the ability to classify the reconstructed data according to the movement performed was assessed. Classifiers were trained to predict movement type from kinematic data using the full training data. The accuracy of the predictions generated by reconstructed kinematic data from a set of features was then assessed. This provides a measure of whether the features that distinguish between movement classes are preserved in the latent features.

A number of linear classification models were used. These included logistic (softmax) regression, support vector machines, stochastic gradient descent classifier, naïve bayes classifiers and random forest classifiers. Hyperparameter optimization was performed via a parameter search using five-fold cross-validation.

The impact of the number of features used on the classification accuracy obtained was assessed by measuring the accuracy achieved for data reconstructed from a range of feature sets from one to 21 using a model trained to classify movements using the full test data.

Movement interpolation

Reconstruction of previously unseen movements was performed by removing all instances of a movement type from the data. Models were then trained on all other movements. This then produced a set of latent features learned on the subset of data used for training that captures the structure of those movements. The ability to reconstruct the unseen movement was then measured. This was performed by providing the previously unseen movement as an input, representing it using the latent features learned on other movements, then reconstructing the original kinematic data using a decoder trained on the other movements. This assesses whether the learned latent features capture a general structure of hand movements that is able to represent and reconstruct new movements outside of the training set, indicating that the method is capturing the true structure of hand kinematics rather than statistical idiosyncrasies of the training movements.

This was repeated for all movements. This allowed simulation of application of the model to novel movements, providing an assessment of its ability to generalize to new movements.

Based on the measurement of each movement's interpolation loss, the movements used were ranked in order of the accuracy with which they could be reconstructed using data from other movements. This provided a measure of which movements contained the most unique movement information, which may allow more accurate characterization of the latent features required to accurately reconstruct movements.

The ability of a reduced set of movement tasks to characterize these latent features and provide a generalizable means of reconstructing all movements was assessed. Models were trained on the training data with a range of movements types removed, from the full set of movements to only the top-ranked movement. The ability of these models trained on limited task sets to reconstruct the full set of all tasks in the test set was then measured. This provided an estimate of the number of training tasks required to identify features that reliably generalize to interpolate all other movements.

Assessing generalization

The ability of the features identified to generalize was assessed in order to evaluate whether movement features were valid across a population or if each individual's movements had a unique structure that prevented generalization. All participants' data were combined and the ability of models trained on this pooled data to reconstruct movements was assessed.

Transfer of these population-based feature sets to individuals was assessed. This simulated a situation where a population model was applied to a new individual, attempting to reconstruct that individual's movements without having had access to any training data from that individual. For each participant, models were trained on the pooled data of all other participants and performance tested on that individual's data. This was repeated for all participants in order to measure the ability to apply population-based features to individual reconstructions.

The ability to learn generalizable features from limited datasets was assessed by measuring reconstruction error and classification accuracy using models trained on limited subsets of the training data, ranging from all training examples (40 examples of each movement) to 10% of the training examples (4 examples of each movement). This was carried out using individualized and transferred models in order to assess the sensitivity of the models to the amount of training data available and the ability to learn useful features from few examples.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

Evaluation of the effect of model used (i.e. principal component analysis or sequential autoencoder) and the training data used (i.e. individualized models using that participant's training data or a population model using all other participants' data) on reconstruction error and classification accuracy was assessed using two-way repeated-measures ANOVA with a pre-specified significance threshold of $p < 0.05$.

Where a significant effect was found on ANOVA, pairwise differences were further evaluated using repeat measures t-tests with a pre-specified significance threshold of $p < 0.05$.

Models were further compared based on the number of features required to achieve accurate reconstructions and the number of training tasks required to learn generalizable features.

Software

All analyses were performed using custom Python scripts.

Numerical computations were performed using NumPy arrays.[45] Linear models were implemented using scikit-learn.[46] Deep learning models were implemented using PyTorch.[47] Statistical models were implemented using statsmodels.[48] Figures were produced using matplotlib[49] and seaborn.[50]