Understanding Motor Control for Assistive Technologies: From Biomechanics to Brain-Machine Interfaces

by

Luis Hernan Cubillos Guzman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Robotics)
in the University of Michigan
2025

Doctoral Committee:

Professor Cynthia Chestek, Co-Chair
Associate Professor Chandramouli Krishnan, Co-Chair
Professor Brent Gillespie
Associate Professor Elliott Rouse

Luis H. Cubillos

lhcubill@umich.edu

ORCID iD: 0000-0001-9876-6204

© Luis H. Cubillos 2025

Dedication

I dedicate this dissertation to my parents, Luis and Catalina, to my sisters, Catalina, Trinidad, Pia, Montserrat, and Magdalena, and to my amazing wife, Sofia.

Acknowledgements

None of this work would be possible without all the people I have had an opportunity to work with, and want to specifically acknowledge all of them here.

First, I want to thank my dissertation committee: Cindy Chestek, Chandramouli Krishnan, Elliott Rouse, and Brent Gillespie. Their support and guidance, most especially during the last few months of my Ph.D., were key to finish the program in the best way possible. I also want to thank Cindy and Mouli, my co-advisors, for their amazing mentorship throughout the Ph.D. It is thanks to both of you that I feel closer to the career goals I had when I started my Ph.D. I also wanted to thank you for not only making the Ph.D. useful and interesting, but also fun: I learned a lot and got a chance to work on some amazing projects; I am very thankful for those opportunities.

Second, I want to thank my lab mates: Tom Augenstein, Peter Washabaugh, Matt Mender, Eric Kennedy, Hisham Temmar, Joey Costello, Dylan Wallace, Joe Letner, Julianna Richie, Maddi Kelberman, Ayo Ward, Jordan Lam, Matteo Ceradini, Olivia Lee, Miranda Copenhaver, Jesse DeMatteo, Jake Joseph, Aren Hite, Alex Vaskov, Jeremy Simmons, Omer Benharush. It is really fun going to work each morning and liking the people at work. I had an amazing time working with all of you and strongly hope our paths cross again in the future.

Third, I want to thank my wife and my family, both the one I was born into as well as the one I chose through my wife Sofia. Your support was amazing throughout the program and kept me going even during the hardest moments.

Finally, I want to thank the research animals, Monkeys N, W, R, K, and H, whose contributions made all this research possible and remind us of the responsibility we have to use their sacrifice to enable discoveries that benefit the lives of all.

Table of Contents

Dedication ii

Acknowledgements iii

List of Tables ix

List of Figures x

List of Appendices xviii

Abstract xix

Chapter 1 Introduction 1

1.1 Human Motor Control 1

1.2 Motor Learning For Refining Motor Control 5

1.3 Prosthetics For Restoration Of Motor Control 6

1.4 Algorithms For Prosthetic Control 8

1.5 Summary Of Thesis 10

Chapter 2 Reliability and Minimal Detectable Change of Stiffness and Other Mechanical Properties of the Ankle Joint in Standing and Walking 13

2.1 Summary 13

2.1.1 Background 13

2.1.2 Research Question 13

2.1.3 Methods 13

2.1.4 Results 14

2.1.5 Significance 14

2.2 Introduction 14

2.3 Methods 16

2.3.1 Participants 16
2.3.2 Perturberator Device 17
2.3.3 Experimental Protocol 19
2.3.4 Impedance Measurements 20
2.3.5 Statistical Analyses 22
2.4 Results 23
2.5 Discussion 27
Chapter 3 Breaking the Barriers to Designing Online Experiments: a Novel Open-Source Platform for Supporting Procedural Skill Learning Experiments 33
3.1 Summary 33
3.1.1 Background 33
3.1.2 Method 33
3.1.3 Results 34
3.1.4 Conclusions 34
3.2 Introduction 34
3.3 Methods 37
3.3.1 Web-based Experimental Platform 37
3.3.2 Feasibility study 41
3.3.3 Data Analysis 43
3.3.4 Data Management and Analyses 46
3.4 Results 47
3.5 Discussion 51
3.6 Conclusion 58
Chapter 4 Exploring Synergies in Brain-Machine Interfaces: Compression vs. Performance 59
4.1 Summary 59
4.2 Introduction 60

4.3 Materials and Methods 62
4.3.1 Experimental Design 62
4.3.2 Statistical Analyses 67
4.4 Results 72
4.4.1 Signal Quality Assessment 72
4.4.2 Relation Between Synergies and Movement 74
4.4.3 Compression and Denoising Experiments 77
4.4.4 Generalization Experiment 80
4.5 Discussion 81
Chapter 5 Exploring the Trade-Off Between Deep-Learning and Explainable Models For Brain-Machine Interfaces 86
5.1 Summary 86
5.2 Introduction 87
5.3 Related Work 89
5.4 Methods 90
5.4.1 Neural Decoders 90
5.4.2 Task and Data Acquisition 93
5.4.3 Decoder Training and Testing 94
5.4.4 Performance Metrics 95
5.4.5 KalmanNet as a Non-linear Trust System 96
5.4.6 Robustness to Real-World Operation 97
5.5 Results 98
5.5.1 Offline and Online Performance 98
5.5.2 Modulating Trust between Dynamics and Observations 101
5.5.3 Robustness to Noise and Change in Task Context 102
5.6 Discussion 104

5.7 Acknowledgements 107
Chapter 6 Discussion 108
6.1 Implications for Motor Control and Motor Learning 108
6.2 Implications for Assistive Technologies 110
6.3 Future Directions 112
Appendices 116
Bibliography 140

List of Tables

Table 2.1: Descriptive statistics for the impedance parameters while standing and walking for the “ankle+robot” case. The last column shows the group MDC as a reference for a hypothetical sample size of N=30. 24

Table 2.2: Descriptive statistics for the impedance parameters while standing and walking for the “only ankle” case. The last column shows the group MDC as a reference for a hypothetical sample size of N=30. 25

Table 3.1**:** Descriptive data for the ten learning metrics in the supervised and unsupervised groups. 48

Table 3.2**:** Descriptive data for the ten learning metrics collapsed between groups 49

Table 4.1: Ridge regressions for denoising analyses 71

Table 4.2: Number of active intracortical channels per day and monkey 73

Table 4.3: Average compression ratios across methods and monkeys. 77

Supp. Table A.1: Descriptive statistics for the impedance parameters while standing and walking for the “ankle+robot” case, separated by sex. 120

Supp. Table A.2: Descriptive statistics for the impedance parameters while standing and walking for the “only ankle” case, separated by sex. 121

Supp. Table A.3: Descriptive statistics for the impedance parameters while standing and walking for the “ankle+robot” case, when normalizing by body mass × height. The last column shows the group MDC as a reference for a hypothetical sample size of N=30. 122

Supp. Table A.4: Descriptive statistics for the impedance parameters while standing and walking for the “only ankle” case, when normalizing by body mass × height. The last column shows the group MDC as a reference for a hypothetical sample size of N=30. 123

Supp. Table A.5: Descriptive statistics for the absolute (non-normalized) impedance parameters while standing and walking for both analyzed cases. 124

List of Figures

Figure 2.1: **Experimental protocol and ankle impedance estimation method.** (A) Each session was split into a standing block and multiple walking blocks. The left image shows an example of a standing perturbation, while the right one is of a walking perturbation, both for a right-leg dominant participant. The blue platform shown in the picture conveys the perturbation and measures the ground reaction forces. (B) The Test and Retest sessions were replicas of each other and were scheduled with at least one day between them. Each session started with a standing block, where participants received 20 perturbations with a 10 to 15 seconds inter-perturbation delay. During a walking block, 20 perturbations were applied: 10 during the early stance and 10 more during the late stance. For each walking trial, there was a 50% chance of applying a perturbation, which means that not all trials contain a perturbation (represented by k unperturbed trials in the figure). (C) Description of the method used to estimate impedance from measured ankle angle and torque while walking. First, ankle angle and torque (with and without the robot's inertia) are extracted from perturbed and unperturbed trials; the platform angle for a sample perturbation is shown in red (left). Then, the ensemble trajectory of the unperturbed trials is subtracted from each perturbed trial trajectory to view the perturbation-only response, and a 100ms window around the perturbation is extracted (10ms before perturbation, 90ms after perturbation; middle part of the figure). The average subtracted angle (with the offset removed) is then numerically differentiated (Scheid, 1968) to extract velocity and acceleration. Finally, angle, velocity, and acceleration are used to fit a second-order model to the average subtracted torque response (with the offset removed) and find the impedance parameters (right). 18

Figure 2.2: **Correlation and Bland-Altman plots for the stiffness and viscosity parameters between sessions for standing and walking**. (A)-(D) show the correlation plots between the Test and Retest sessions for stiffness (first column) and viscosity (second column) during standing (first row) and walking (second row). Blue circles represent values from individual values in the "only ankle" case. Orange squares represent individual values in the "ankle + robot" case. The blue and orange segmented lines represent a linear fit to the data from the "only ankle" and "ankle + robot" cases, respectively, while the black segmented lines show a line of unit slope and zero intercept for reference. The blue and orange text represents the Pearson correlation coefficient between Test and Retest for each corresponding case. (E)-(H) show the Bland-Altman plots for the Test and Retest sessions for stiffness (first column) and viscosity (second column) during standing (first row) and walking (second row). Blue circles and orange squares represent individual values for the "only ankle" and "ankle + robot" cases, respectively. The x-axis on each plot corresponds to the mean between the Test and Retest sessions for each individual value and the y-axis represents the difference between the sessions. The solid blue and orange lines show the mean difference for each case, while the dotted lines represent the 95% confidence interval (1.96 * SD). Stiffness and viscosity showed good correlation and agreement, being close to the dashed black line in the correlation plots in

most cases and remaining inside the 95% confidence interval in the Bland-Altman plots. Note for interpretation of the references to color in the figure legend, please refer to the online version of this article. 26

Figure 3.1: **Web application flowchart**. A schematic of the webpage flow as seen by the two types of users: researchers (top) and participants (bottom). Researchers log in to the platform and are then presented with a profile page. There, they can create either a study or an experiment, manage unpublished and published studies, and test experiments (mock examples shown). Participants enter the study, experiment, and group codes given to them by the researchers, and then are redirected to the corresponding experiment. 39

Figure 3.2: **Schematic of experimental protocol and data processing**. (A) A schematic of the experimental protocol. Participants practiced a finger tapping task where they first learned to type the sequence '41324' as fast and as accurately as possible with their non-dominant, left hand. After a one-minute rest following the completion of the task with their left hand, they then learned to type the sequence '70897' as fast and as accurately as possible with their right hand. Participants received 36 blocks of training for each hand with each block consisting of 10 seconds of practice and 10 seconds of rest. (B) A schematic for the computation of micro-scale learning (micro-offline and micro-online learning). Micro-offline learning quantified the learning that occurred during short periods of rest, and micro-online learning quantified the learning that occurred during short periods of practice. For block i, micro-offline learning is defined as the difference in tapping speed (keypresses/second) between the first trial of block i and the last trial of block i-1. Micro-online learning is defined as the difference between the last and first trials of block i. Total micro-learning at that block is the sum of both quantities. (C) A schematic of the interlimb transfer of motor skill learning metrics. Early transfer is calculated as the difference in the mean trial tapping speed between the first block of the right and left hand. Late transfer is calculated as the difference in mean trial tapping speed between the last block of the right and left hand. 44

Figure 3.3: **Tapping speed performance**. A scatterplot showing the mean tapping speed (keypresses/second) of the supervised and unsupervised groups in each block of practice for the (A) left and (B) right hand of the participants. The shaded region indicates the standard error of the mean (SEM). Note that the performance on the finger tapping task was similar between the supervised and unsupervised groups for both hands. Also, note that the tapping speed on the initial block of the right (i.e., transfer) hand was higher than the left (i.e., training) hand, indicating the presence of interlimb transfer of procedural skill learning. 50

Figure 3.4: **Classification using Linear Discriminant Analysis (LDA) model:** Scatterplot showing the projection of the ten features described in Table 3.1 and Table 3.2 to the first two discriminant coordinates of a linear discriminant analysis (LDA) model. Each marker represents a single participant, and the color represents the group they belong to. No clear separation can be observed between the two groups, and no significant difference in classification accuracy was found between the LDA classifier and a random classifier (mean $p = 0.55$). Note that for visualization purposes only the first two discriminant coordinates have been plotted. 50

Figure 3.5: **Micro-scale learning**. A scatterplot showing the mean micro-scale (micro-offline, micro-online, and total) early learning (first 11 blocks) for the **(A)** left and the **(B)** right hand across all participants. Micro-offline learning was defined as the difference in tapping speed (keypresses/second) between the first trial of a practice block and the last trial of the previous block and quantified learning that occurred during short periods of rest. Early micro-offline learning was computed as the sum of the micro-offline learning for the first 11 blocks. Micro-online learning was defined as the difference in tapping speed between the last and the first trial of a practice block and quantified learning that occurred during short periods of practice. Early micro-online learning was computed as the sum of the micro-online learning for the first 11 blocks. Total micro-learning was defined as the sum of the micro-offline and micro-online learning and quantified the learning between blocks. Early total micro-learning was calculated as the sum of total micro-learning in the first 11 blocks and represented all changes in tapping speed during early learning. The shaded region represents the standard error of the mean (SEM). The horizontal dashed line shows the zero value. Note that micro-online learning was negative for most of the blocks, indicating that participants had higher tapping speeds at the beginning than at the end of each block. Traces showing the mean tapping speed for the **(C)** left and **(D)** right hand across all participants for the ten seconds in each block during early learning. The shaded region represents the SEM. Note that this plot was created only for visualization purposes and the details for creating this plot are provided in Appendix B:. 51

Figure 3.6**: Distribution of micro-scale learning:** Violin plots showing the distributions of early micro-offline learning, early micro-online learning, and early total micro-learning for the **(left)** left and **(right)** right hand across all participants. Early micro-learning (micro-offline, micro-online, and total) was computed as the sum of the values observed in the first 11 blocks. Note that the white circles indicate the median values, the filled circles indicate individual data points, and the width of the shaded region represents the approximate frequency of data points in each region. 52

Figure 4.1: **Monkey task and data processing pipeline**. (A) Monkeys N and W were trained to do a 2-DoF dexterous finger task, in which they had to move the index and the middle-ring-small (MRS) fingers independently to targets shown on the screen (center diagram). While they did the task, we recorded finger position and velocity (top right plot), spiking band power of the brain channels (top left plot) and, for Monkey N, muscle activity (bottom left plot). (B) Diagram showing the data processing pipeline for the muscle (left) and brain (right) data. The 7 bipolar muscle channels are processed to extract envelopes and then normalized to the peak activation from each day. Then, we applied PCA and NMF, resulting in a set of muscle synergy activations (top plot) and synergy weights (bottom plot). The 96 brain channels are filtered to extract only those with at least 1Hz firing rate on average, and then transformed using the SBP in 20ms bins. PCA, NMF, and dPCA are then applied, which results in a set of brain synergy activations. **Abbreviations**: MRS = middle-ring-small fingers; EMG = electromyography; PCA = principal component analysis; NMF = non-negative matrix factorization; dPCA = demixed PCA; SBP = spiking band power; FDP = flexor digitorum profundus; EDC = extensor digitorum communis; FCR = flexor carpi radialis; ECRB = extensor carpi radialis brevis; EIP = extensor indicis propius. 64

Figure 4.2: **Assessment of the quality of physiological signals**. (A) Monkey N's raw and processed EMG, together with MRS finger velocity for two example trials. Gray vertical bars

represent the removed juicer artifact *(see Methods)* at the beginning of each trial. Top plot shows FDP, a muscle targeting MRS flexion, and the bottom shows EDC, a muscle targeting MRS extension. Vertical gray line shows scale of 100uV for the raw EMG and horizontal black line shows scale for 500ms. (B) Example traces of Monkey N's processed FDP EMG predicted from brain activity (top) and finger velocity predicted from brain activity (middle) and processed EMG (bottom). Bar plot on the right shows aggregated correlations corresponding to three days of Monkey N data. *Black* dots *in the left bar* represent the predictions for *each* muscle channel (*seven dots for each day*). *Bar plots in middle and right bar represent the velocity predictions for each finger group (two dots for each day).* (C) Accuracy of classifying targets using EMG (muscle symbol) and brain activity (brain symbol) for both monkeys across days. Segmented line shows chance level. Aggregated results correspond to three days of Monkey N and two days of Monkey W. (D) Same as in B but for Monkey W. Trace shows example of MRS velocity and the prediction when using the brain data. Bar plot shows aggregated results across 2 days of *M*onkey W data. **Abbreviations:** FDP = flexor digitorum profundus; EDC = extensor digitorum communis; MRS = middle-ring-small fingers; Vel. = velocity; EMG = electromyography. 74

Figure 4.3: **Analysis of brain and muscle synergy relationships for monkey N**. (A) Muscle synergy weights extracted using PCA (top bar plot in each window) and NMF (bottom bar plot) across days. Synergy weights were consistent across days and methods and revealed a structure closely related to the task. (B) Barplot showing the average across days of the maximum absolute correlation coefficient between all muscle channels and all brain channels with the four kinematic outputs: index and MRS position, index and MRS velocity. The brain channels had a weaker direct relationship with the kinematic outputs, which hinders an interpretation of the brain synergies similar to that shown in A. (C) Average of the absolute correlation coefficient between all possible pairs of brain and muscle synergy activations, broken down by the method(s) used. The *black* segmented line shows the average of the absolute correlation coefficient between all *pairs of* brain and muscle channels, while the shading represents the standard error of the mean (SEM). All error bars represent the SEM as well. **Abbreviations**: PCA = principal component analysis; NMF = non-negative matrix factorization; FCR = flexor carpi radialis; FDP = flexor digitorum profundus; FDPd = FDP distal; FDPp = FDP proximal; ECRB = extensor carpi radialis brevis; EIP = extensor indicis propius; EDC = extensor digitorum communis; Idx = index finger; MRS = middle-ring-small fingers; Pos. = position; Vel. = velocity; Syn. = synergy; Corr. Coef. = Pearson correlation coefficient; a.u. = arbitrary units. 76

Figure 4.4: **Compression and denoising analyses**. (A) Diagram explaining the connections that were tested. Each line represents a different model. For example, the pink line represents a model that was trained with brain synergy activations as the input, and with muscle activity as the output. The three icons represent the brain channels, the muscle activity, and the finger velocity, respectively. The two black lines represent the baselines, shown in black in the rest of the plots. (B, top) Correlation coefficient of the prediction of muscle activity from brain synergies (left) and from the brain channels through muscle synergies (right). The x axis represents what percent of the total dimensions were used to make the predictions. The black lines represent the result of using all brain channels to predict all muscle activity. The colors of the icons are as in A. The vertical lines represent the percent of total dimensions that are needed for each method to achieve 95% of the maximum correlation. The shading for each line

represents the standard error of the mean across days. (B, bottom) Correlation coefficient of the prediction of finger velocity from brain synergies (left) and from muscle synergies (right). (C) Average variance accounted for as a function of the percentage of total brain (left) and muscle (right) channels for monkey N. (D) *Sample h*eatmap showing, *for day 1,* the correlation of predicting muscle activity through brain and muscle synergies, normalized to the result of predicting muscle activity directly from brain channels. A value of 100 would mean that it matches the *baseline* perfectly. P represents PCA, N represents NMF, and d represents dPCA. (E) Same as B and C but for Monkey W. **Abbreviations**: PCA = principal component analysis; NMF = non-negative matrix factorization; dPCA = demixed PCA; VAF = variance accounted for; Syn. = synergy; Vel. = velocity..... 79

Figure 4.5: **Generalization analyses for Monkey N**. (A) Diagram showing how the on-context and off-context models differ: on-context models are trained on each non-normal context (spring, wrist, spring+wrist) and tested on the same context. Off-context models are trained on the normal context and then tested on each of the non-normal contexts. (B) Average across non-normal contexts of the Pearson correlation coefficient of the prediction of muscle activity from all numbers of brain synergies up to the total dimension of the brain data. The two black lines represent the baselines for on-context and off-context, when predicting the muscle activity from the full brain data. (C) Same as in B but for the prediction of muscle activity from the full brain data and through muscle synergies (see Methods). (D) Same as in B and C but now showing the prediction of finger velocity from brain synergies. (E) Same as in B, C, and D, but now for the prediction of finger velocity from muscle synergies. (F) Heatmap for the spring+wrist context, like the one shown in Figure 4.4D, showing the correlation coefficient of the prediction of muscle activity from brain synergies and through muscle synergies, normalized to the off-context baseline prediction. A value over 100 means a better prediction than the baseline off-context model that uses all brain data and predicts muscle activity directly. **Abbreviations**: PCA = principal component analysis; NMF = non-negative matrix factorization; dPCA = demixed PCA; Syn. = synergy. 81

Figure 5.1: **Task and Neural Decoders.** (A) We trained a monkey to do a 2-DoF finger task (shown on screen) while brain activity and finger kinematics (index and middle-ring-small (MRS) traces shown below) are recorded. The monkey can do the task in hand control, using his hand, or in brain control. (B) In brain control, the SBP of each brain channel is extracted and binned every 50 milliseconds. Each neural decoder takes in a bin and predicts position and velocity, or only velocity. (C) The KF and KalmanNet differ in how they compute the Kalman gain: the KF uses the equation shown, while KalmanNet uses a set of RNNs. 94

Figure 5.2: **Offline Performance.** (A) Traces of ground truth index position and velocity (in blue) versus the predictions from each neural decoder. Note that tcFNN only predicts finger velocity. (B) Velocity (left) and position (right) performance in terms of correlation and MSE for each neural decoder. Square markers and error bars denote the mean and the standard error of the mean, respectively. Tested across $n=13$ days from both monkeys..... 99

Figure 5.3: **Online Performance.** (A) Traces of the positions of the index (blue) and middle-ring-small (yellow) fingers during online control, across neural decoders. Blue-filled boxes represent index targets in the flexion-extension range, while yellow-filled boxes represent MRS targets. Green-outlined boxes represent successful trials, while red-outlined boxes show

failed trials. (B) Online metrics of performance. Throughput, trial times, and path efficiencies are normalized to the corresponding KalmanNet value for each day. Tested across $n=5$ days from Monkey N..... 100

Figure 5.4: **KalmanNet as a Non-linear Trust System**. (A) Offline trace showing the norm of the Kalman gain for KalmanNet (in green) and for the KF (in orange), together with the predicted velocity for KalmanNet (black). The inset shows the first five seconds of trials, during which KF's Kalman gain converges to a fixed value~\cite[Ch. 4]{durbin2012time}. (B) Average correlations between Kalman gain and velocity for KalmanNet, HKF, and KF, for offline ($n=13$ days, across both monkeys) and online ($n=5$ days, only Monkey N) trials. (C) Online trace of the norm of the Kalman gain for KalmanNet (green), together with the predicted velocity (black). (D) Offline velocity prediction performance comparison between KF, HKF, and KalmanNet, in terms of correlation coefficient (left) and MSE (right). Arrows represent the addition of a heteroscedastic process noise to transform the KF into HKF. Tested across $n=13$ days and both monkeys..... 101

Figure 5.5: **Robustness to Injected Noise**. (A) Offline traces across neural decoders as compared to the ground truth index velocity (black). Grey columns represent noise injections of 5 times the standard deviation of the training dataset distribution, for a duration of one time bin. (B) Change in normalized velocity MSE as the noise magnitude changes, for a fixed noise duration of 3 time bins. (C) Change in normalized velocity MSE as the noise duration changes, for a fixed noise magnitude of 1x the training dataset distribution's standard deviation. Tested on 13 days for Monkeys N and W.} 103

Figure 5.6: **Generalization Across Task Contexts**. (A) Illustration of how off-context and on-context decoders are trained. Monkeys performed a variation of the task that included a change in wrist angle and a spring that resisted flexion. Off-context decoders were trained on a normal set of trials from the same day and then tested on the last 20% of the trials of the spring+wrist context. On-context decoders were trained and tested on the same spring+wrist context. (B) Change in velocity MSE between on and off-context decoders for each decoder. Tested across n=6 days of Monkey N. 104

Supp. Figure A.1: **Correlation and Bland-Altman plots for the inertia parameter between sessions for standing and walking**. (A)-(D) show the correlation plots between the Test and Retest sessions for the “only ankle” case (first column) and “ankle + robot” case (second column) during standing (first row) and walking (second row). Blue circles represent values from individual values in the “only ankle” case. Orange squares represent individual values in the “ankle + robot” case. The blue and orange segmented lines represent a linear fit to the data from the “only ankle” and “ankle + robot” cases, respectively, while the black segmented lines show a line of unit slope and zero intercept for reference. The blue and orange text represents the Pearson correlation coefficient between Test and Retest for each corresponding case. (E)-(H) show the Bland-Altman plots for the Test and Retest sessions for the “only ankle” case (first column) and “ankle + robot” case (second column) during standing (first row) and walking (second row). Blue circles and orange squares represent individual values for the “only ankle” and “ankle + robot” cases, respectively. The x-axis on each plot corresponds to the mean between the Test and Retest sessions for each individual value and the y-axis represents the difference between the sessions. The solid blue and orange lines show the mean

difference for each case, while the dotted lines represent the 95% confidence interval (1.96 * SD). The inertia showed good correlation for all but the “only ankle” during walking and was close to the dashed black lines. Bland-Altman plots show good agreement with no apparent systematic errors. 117

Supp. Figure A.2: **Ankle torque during standing**. Left: ankle torque due to perturbation while standing for a representative subject and session. Each colored trace represents a single perturbed trial, which had its offset (the value right before perturbation) removed. The blue line shows the average of all the 100 ms mean subtracted unperturbed segments before each perturbation. The perturbed torque curves were of greater amplitudes and consistent shapes. Right: comparison between the absolute (non-mean subtracted) ankle torques for 300ms segments of unperturbed trials (blue, shaded region represents standard deviation) and perturbed trials (red, shaded region represents standard deviation). The first green dashed line indicates the moment when perturbation happened and the analysis window started, while the second green dashed line indicates when the analysis window ended. 118

Supp. Figure A.3: **Perturbation characteristics during walking**. (top, left): Histogram of perturbation timings during walking across subjects and sessions. The vertical black dashed lines represent the average perturbation timing for the early (left line) and late (right line) stance perturbations. (top, right): exemplary ensemble average of early stance perturbed (orange) and unperturbed (blue) ankle torque during walking trials, starting at the average time of perturbation. The perturbed curve shows a more systematic shape, which corresponds to the torque response to the perturbation. (bottom, left): Histogram of ankle angles (in degrees) at perturbation onset during walking, for early (blue) and late (orange) stance. The average ankle angles at each stage are shown with vertical black dashed lines. (bottom, right): Average and standard deviations of the range of ankle angles across all participants for both sessions (range = max. angle in a session – min. angle in a session), separated by when the perturbation happened (early stance: blue, late stance: yellow). 119

Supp. Figure B.1: **Micro-scale learning across groups**: Scatterplot showing the mean micro-scale (micro-offline, micro-online, and total) early learning (first 11 blocks) for the left (A and C), and right (B and D) hand, and for the supervised and unsupervised groups, respectively. Description of the plots is the same as in Figure 3.4 125

Supp. Figure B.2: **Distribution of micro-scale learning across groups**: violin plots showing the distribution of early micro-offline learning, early micro-online learning, and early total micro-learning for the (A and C) left and (B and D) right hand, and for the supervised and unsupervised groups, respectively. Early micro-learning (micro-offline, micro-online, and total) was computed as the sum of the values observed in the first 11 blocks. Note that the white circles indicate the median values, the filled circles indicate individual data points, and the width of the shaded region represents the approximate frequency of data points in each region. 126

Supp. Figure C.1: **Heatmap of muscle activity prediction from brain synergies to muscle synergies for day 2** (left) and 3 (right). Each cell value corresponds to the correlation of predicting muscle activity through brain and muscle synergies, normalized to the result of predicting muscle activity directly from brain channels. A value of 100 would mean that it

matches the baseline perfectly. P represents PCA, N represents NMF, and d represents dPCA. 129

Supp. Figure C.2: **Heatmap of muscle activity prediction from brain synergies to muscle synergies the spring** (left) and wrist (right) contexts on day 1. Each cell value corresponds to the correlation of predicting muscle activity through brain and muscle synergies, normalized to the result of the off-context muscle activity prediction directly from brain channels. A value of 100 would mean that it matches the baseline perfectly. P represents PCA, N represents NMF, and d represents dPCA. 130

Supp. Figure D.1: **KalmanNet architecture**. Diagram of the components of the KalmanNet network. It consists of three GRUs plus seven linear + ReLU layers that try to model the normal way of computing the Kalman gain (Revach et al., 2022). F1 through F4 correspond to the input features from the previous equations. 134

Supp. Figure D.2: **Sensitivity analysis of sequence length during training**. Offline velocity correlation (left) and MSE (right) for KalmanNet, under different sequence lengths employed during training. The horizontal axis represents the number of 50ms bins; the one used throughout corresponds to 60, or equivalently, three seconds. Computed across all n=13 days for both monkeys. 136

Supp. Figure D.3: **Resistance to noise injection**. (A) Offline velocity MSE for KalmanNet (green) and LSTM (brown) across n=13 days for both monkeys, with noise values closer to those present in the training data. A noise of zero magnitude is equivalent to not adding noise (i.e., baseline shown in Figure 5.2). (B) Full product of normalized velocity MSE across models for all values of noise magnitude and duration. The logarithmic color bar on the right represents the MSE value for each combination of noise magnitude and duration, normalized to each model's baseline performance (without noise). 139

List of Appendices

Appendix A: Supplement to Chapter 2 117

Appendix B: Supplement to Chapter 3 125

B.1 Supplementary Figures 125

B.2 Variables Used 126

B.3 Visualization Plot 128

Appendix C: Supplement to Chapter 4 129

Appendix D: Supplement to Chapter 5 131

D.1 Kalman Filter 131

D.2 KalmanNet 133

D.3 Heteroscedastic Kalman Filter 136

D.4 tcFNN 137

D.5 LSTM 138

D.6 Supplemental Figures 139

Abstract

Millions of people in the world suffer from disorders that affect their motor control, including conditions like spinal cord injuries, strokes, and neurodegenerative diseases. These impairments can drastically reduce their mobility and independence, creating a need for advanced assistive technologies. This dissertation addresses these challenges by exploring how humans move and learn new motor skills, and how these insights may inform better assistive technologies, especially through brain-machine interfaces.

We first focus on the biomechanics of the ankle joint, studying the reliability of its stiffness, viscosity, and inertia while standing and walking. The results show that stiffness and viscosity can be reliably estimated and may be valuable as a clinical metric as well as a goal for biomimetic prosthetic control.

Then, we show the development of a new open-source web application for motor learning experiment design and show its validation for replicating previous studies, as well as for investigating how motor learning is transferred between limbs. Our results show that even short periods of rest can greatly improve motor skill learning as well as the existence of interlimb transfer learning.

Finally, we apply these insights in the context of brain-machine interfaces, assistive devices that have helped restore motor and speech control by decoding signals from the brain. We first evaluate whether using latent dimensions, called brain and muscle synergies, can help in decoding brain and muscle data by aiding in compression, denoising, and generalizing to other task contexts. We found improvements in data compression, but not in denoising or generalization. Second, we evaluate the trade-offs of using an interpretable algorithmic approach that combined a principled approach with deep-learning for decoding finger movements from brain data. We found that explainability may not need to be sacrificed for good performance, and that an

interpretable algorithm can provide insights into motor control as well as future avenues for model improvement.

Overall, by integrating these distinct but related areas, this dissertation contributes to a comprehensive understanding of human motor control and how to improve it. Together, these contributions provide a foundation for assistive technologies that can restore mobility, enhance independence, and improve quality of life for individuals living with motor impairments.

Chapter 1 Introduction

Understanding motor control and its refinement is essential for developing effective strategies to restore lost movement functionality through assistive technologies. Section 1.1 explores the fundamentals of human motor control, focusing on how the nervous system coordinates and executes movements through various levels of abstraction. Building on this, Section 1.2 discusses motor learning as a process for refining motor control through motor skill learning, and what common experimental paradigms for studying it exist. The advancements in prosthetics for restoring motor control are introduced in Section 1.3, highlighting how prosthetic devices can replace or augment lost limb functionality and how neuroprostheses can help augment or replace parts of the nervous system. Section 1.4 goes into algorithms for prosthetic control, making the distinction between linear and non-linear models and highlighting the applications for prosthetic control that they have enabled. Finally, Section 1.5 summarizes the key points of this thesis and outlines its contributions to the field.

1.1 Human Motor Control

Humans can be viewed as highly sophisticated biological robots, capable of perceiving their environment, making decisions, and executing movements through a series of precise, coordinated processes. At the core of this system is the nervous system, which functions like the body's control architecture. The central nervous system (CNS)—composed of the brain and spinal cord—works alongside the peripheral nervous system (PNS), which transmits signals between the CNS and the rest of the body. Neurons, the fundamental units of this system, act like

electrical components, sending information through action potentials across trillions of connections to perform complex computations. For example, even a simple action like flexing a finger starts with the motor cortex, a brain region responsible for planning and initiating movement (Bear et al., 2020; Kandel et al., 2021). Upper motor neurons generate action potentials that travel through the spinal cord and out via nerves, such as the ulnar and median nerves, to reach muscles like the flexor digitorum profundus, a finger flexor muscle (Bear et al., 2020; Lieber, 2002; Mender et al., 2023). These signals trigger biochemical processes that cause muscle fibers to contract, resulting in movement. This feedforward control process mirrors robotic systems, transforming intention into precise motion and showcasing the human body's mechanical and computational design.

This feedforward process can be studied on three levels of abstraction: how the desired movement is computed and represented at the level of the neurons in the motor cortex, how this computation relates to the actual control of muscles, and how this control of muscles relates to kinetic and kinematic outputs. First, regarding the computation of movement. A few decades ago, the prevailing theory of movement computation stated that the activity of a neuron was fully explained by the representation of specific movement parameters, such as movement direction (Butts & Goldman, 2006; Carmena et al., 2003; Lalazar et al., 2016). However, as the number of neurons available for live recording has been increasing (Stevenson & Kording, 2011), a growing body of evidence has started to argue for an interpretation of this activity in the context of dynamical systems (Abbaspourazad et al., 2023; Churchland et al., 2012; Gallego et al., 2017a; Russo et al., 2018). Under this interpretation, the representation of movement parameters in the firing rate of a motor neuron is a symptom of the underlying driving activity, rather than the cause of the movement (Churchland et al., 2012). This principle has motivated studying the

neurons' population activity and the space in which this activity lives, also called the neural manifold (Abbaspourazad et al., 2023; Gallego et al., 2017a). Studying the neural manifold has shown that the computation of movement at the level of the motor cortex has low "tangling", where moments of similar activity patterns lead to similar future activity (Russo et al., 2018; Saxena et al., 2022). Additionally, it has also been shown that this neural manifold is similar across tasks (Gallego et al., 2017a, 2018a), even with different muscle activity requirements, and consistent throughout time (Gallego et al., 2020).

If the activity of the neurons in the motor cortex does not directly represent muscle activity downstream, then how is muscle activity computed? The brain can coordinate the activity of hundreds of muscles within the body to achieve smooth desired movements (Kandel et al., 2021), which makes for a very complex high-dimensional control system. This complexity has led a group of researchers to propose that the brain must lower the dimensionality of the control by controlling functional groups of muscles, called muscle synergies, rather than individual muscles (d'Avella et al., 2003a; Holdefer & Miller, 2002; Overduin et al., 2015a; Tresch & Jarc, 2009). This is still a highly debated issue, but there has been some evidence of a neural basis for these muscle synergies at the level of the motor cortex (Kutch & Valero-Cuevas, 2012a; Overduin et al., 2015a) as well as in the spinal cord (Takei et al., 2017). Researchers who oppose this theory argue that muscle synergies found in other studies represent task constraints rather than an internal brain representation (Kurtzer et al., 2006; Kutch et al., 2008; Valero-Cuevas et al., 2009).

The final level of abstraction in the process of generating movement is how the output kinematics and kinetics are modulated. Evidence has shown that when moving, humans not only modulate the movement trajectory but also the mechanical properties of each joint (Gomi &

Kawato, 1997; Lee et al., 2016), generally referred to as the joint impedance. This modulation of joint impedance has been demonstrated in multiple experiments ranging from walking (Lee et al., 2016; Shorter et al., 2021) to arm-reaching (Popescu et al., 2003; Selen et al., 2006). By modulating joint impedance, humans can increase stability against perturbations in an uncertain environment. For example, we increase the stiffness of the shoulder, elbow, and wrist when receiving a tennis serve (Tsuji et al., 2004), as we are unsure about the speed of the incoming ball, and an increased stiffness prevents the arm from going back unintended and then allows to return the serve. The mechanical impedance of joints is modulated by coordinating the activation of agonist and antagonist muscles (Joshi et al., 2022; Popescu et al., 2003).

A key element of human motor control, however, is that we have access to sensory feedback, in the form of visual, proprioceptive, and tactile, to improve accuracy and reliability in our movements (Kandel et al., 2021). The importance of this feedback has been widely studied first, by looking at people without it (Cole & Waterman, 1995; Prechtl et al., 2001) due to some permanent or temporary condition, and second, by adding sensory feedback to, for example, prosthetic devices, and looking at the performance with or without the feedback (Cimolato & Raspopovic, 2024; Flesher et al., 2021). People without access to some forms of sensory feedback have shown lower accuracy in movements (Brown et al., 2013; Flesher et al., 2021; Gonzalez et al., 2021; Saunders & Vijayakumar, 2011) and force exerted (Brown et al., 2013). An important property of the sensory feedback, however, is that there is an inherent delay in the order of 100 milliseconds (Kandel et al., 2021; Rohde & Ernst, 2016; Stepan, 2009) between the moment the sensor—i.e., visual, proprioceptive, or tactile—gets stimulated and the moment this fact is processed by the brain cortex. This delay suggests that, for fast movements, the brain must have some sort of internal model that, given a desired state, can compute the required muscle

activations (Desmurget & Grafton, 2000; Kandel et al., 2021). Sensory feedback, in those cases, may be used to perform small adjustments, required in cases when the muscle is, for example, more fatigued than expected or when the internal model got the prediction wrong, such as when lifting a heavy can that we thought was empty (Meyer et al., 1988; Milner, 1992; Plamondon & Alimi, 1997).

1.2 Motor Learning For Refining Motor Control

A key enabler for the complexity of human motor control is that it is not static, but rather evolves over time, both in the short term, through motor adaptation (Izawa et al., 2008; Krakauer et al., 2019; Wei & Körding, 2009) and in the long-term, through motor skill learning (Higgins, 1991; Krakauer, 2006; Newell, 1991). Motor adaptation refers to the short-term changes we make to our internal model when given sensory feedback that conflicts with the expected result. Motor skill learning, on the other hand, is a longer-term process that involves either creating a new internal model or the refinement of an existing one (Kandel et al., 2021; Krakauer et al., 2019), for more permanent adjustments.

Most motor skills can be broken down into a series of actions that must be done, either discretely (such as when typing a word on a computer) or continuously (such as the sequence of motor activations needed to hit a backhand in tennis) (Krakauer et al., 2019). Studying continuous sequences is challenging, as it is difficult to control for the many ways of doing the same task across participants (Krakauer et al., 2019). Thus, researchers in the motor skill learning field generally focus on discrete sequence learning, with simple sequences and serial reaction time tasks (SRTT) being two of the most common experimental paradigms used. During simple sequence tasks (Kami et al., 1995; Karni et al., 1998), participants are instructed to complete a sequence of actions as fast and accurately as possible. A common example of a

simple sequence task is the sequential finger-tapping task, in which participants must type a sequence of characters using a keyboard as fast and as accurately as possible (Bönstrup et al., 2019, 2020; Korzeczek et al., 2020; Van Der Werf et al., 2009; Witt et al., 2010).

Most experiments that study motor learning, however, are done in controlled lab environments (Tsay et al., 2021) and many times are underpowered for the hypothesis under testing (Lohse et al., 2016). These weaknesses have driven researchers in the field to conduct online studies (Anglada-Tort et al., 2022; Bönstrup et al., 2020; Listman et al., 2021; Tsay et al., 2021), which can typically recruit orders of magnitude more participants, as well as participants from potentially more diverse populations (Paolacci & Chandler, 2014). Tools such as PsychoPy (Peirce, 2007), Gorilla (Anwyl-Irvine et al., 2020), lab.js (Henninger et al., 2022), and PsyToolkit (Stoet, 2017) are some of the key enablers of this transition to online studies.

1.3 Prosthetics For Restoration Of Motor Control

Neurological conditions such as spinal cord injury (SCI), limb loss, and amyotrophic lateral sclerosis (ALS) significantly impair motor control, leading to disability worldwide. In 2019, approximately 20.6 million individuals globally were living with SCI, with an incidence of about 0.9 million new cases that year (Safdarian et al., 2023). In the United States alone, the prevalence of traumatic SCI is estimated at around 302,000 individuals, with approximately 18,000 new cases annually (Jain et al., 2015; Lasfargues et al., 1995). A spinal cord injury severs or damages the connection between the brain and the spinal cord, which affects downstream motor commands, as well as upstream sensory information. ALS, a progressive neurodegenerative disease, affects around 9.1 per 100,000 people in the United States, leading to the degeneration of motor neurons and resulting in loss of voluntary muscle control (Mehta et al., 2023). Limb loss, often due to trauma or medical conditions like diabetes, significantly impairs

motor function and affects millions worldwide. In the United States, approximately 185,000 amputations occur each year, with nearly 2 million people living with limb loss (Caruso & Harrington, 2024). These conditions greatly diminish individuals' quality of life, which highlights the need for solutions that can restore or improve their motor control.

Prosthetic devices for individuals with limb loss have advanced significantly over the past few decades. For lower-limb amputees, modern prostheses have helped replace the ankle (Azocar et al., 2020; Gabert et al., 2020) and the knee function (Azocar et al., 2020; Lenzi et al., 2018). These prostheses have allowed users not only to do level ground walking but also climb stairs (Au et al., 2008; Azocar et al., 2020) and move up and down across ramps (Azocar et al., 2020; Hargrove et al., 2015), among many other tasks. Advancements in controllers and material science have also enabled lower-limb prostheses that can emulate the mechanical impedance of human joints, thus providing more natural gait patterns (Best et al., 2023; Gabert et al., 2020). Upper-limb prosthetics have shown similar progress (Marinelli et al., 2023). Powered prostheses have been developed for trans-radial (e.g., DEKA “LUKE” arm and the Ability Hand, by Psyonic) and trans-humeral (e.g., Dynamic arm, from Ottobock, and the Espire elbow, from Steeper Inc.) amputations, restoring full arm movements to users. These prostheses have shown increasing degrees of freedom (DoF) for control, from individual finger control (Hotson et al., 2016; Nguyen et al., 2021), to multi-DoF wrist and elbow joints (Mereu et al., 2021; Zbinden et al., 2023). Additionally, these prosthetic devices have shown an increase in the number and type of sensors they contain, allowing the restoration of sensation for amputees, through techniques such as non-invasive sensory substitution (Bark et al., 2014; Markovic et al., 2020), intracortical micro-stimulation (ICMS) (Flesher et al., 2016; Greenspon et al., 2024), or regenerative peripheral nerve interfaces (RPNIs) (Gonzalez et al., 2022; Vu et al., 2020).

A subfield of prosthetic devices, called neuroprosthetics, addresses conditions where there is not necessarily a limb to replace, such as ALS or SCI. Brain-machine interfaces (BMIs) are one group of neuroprosthetics that has seen big improvements in the past few decades. BMIs work by reading from the nervous system, either directly from the brain or at some other point further downstream, passing those signals through some sort of decoding algorithm, and then generating useful control commands (Carmena et al., 2003; S.-P. Kim et al., 2008; Wu et al., 2004). Reading from the nervous system takes many forms depending on the level of invasiveness and where in the nervous system we want to record from. Electromyography (EMG) recordings, for example, measure the electrical activity of the muscles, either by placing electrodes directly on the muscle fibers (Kandel et al., 2021; Mender et al., 2023) or on the surface of the skin (Merletti et al., 2010). Another recording paradigm that has become more common with improvements in electrode design and surgical techniques is intracortical recordings (Musk & Neuralink, 2019; Nordhausen et al., 1994), where microelectrodes capable of recording the activity of individual neurons are inserted into the brain cortex. Intracortical recordings have enabled incredible applications for restoring motor control, from the control of a simple cursor in a screen (S.-P. Kim et al., 2008), controlling a robotic arm (Collinger, Wodlinger, et al., 2013) or their own arm (Ajiboye et al., 2017), to restoring communication capabilities through handwriting (Willett et al., 2021) and speech (Card et al., 2024; Willett et al., 2023).

1.4 Algorithms For Prosthetic Control

The different physiological signals, such as EMG or intracortical recordings, which can be used for prosthetic control, have a complex relationship with movement. Even a simple movement, such as reaching for a cup, can be executed in many different ways by activating

different groups of muscles or by changing postures, which changes the biomechanic constraints of muscle activation (Lieber, 2002; Mender et al., 2024). Additionally, muscle activity is produced through the sequential activation of many neurons starting in the motor cortex (Bear et al., 2020; Kandel et al., 2021), which makes the cortical activity even harder to relate to movement.

To address this challenge, and generate useful control commands from physiological signals, researchers have used different models which can be broadly classified into two: linear and non-linear. Linear models assume a linear relationship between the inputs (i.e., physiological signals) and the outputs (i.e., control commands): when the inputs increase or decrease, the outputs change at a constant rate in a way that the relationship between them can be represented by a straight line in a plot. Linear models that have been widely used in the context of prosthetics are the generalized linear model (GLM) (Collinger, Wodlinger, et al., 2013; Flesher et al., 2021) and the Kalman filter (Gilja et al., 2012; Nason et al., 2021a; Wu et al., 2004). Although linear models assume a possibly wrong view of the relationship between inputs and outputs (Abbaspourazad et al., 2023; Sani et al., 2021), they rely on few assumptions and are simple, which makes them work reliably on user-in-the-loop applications (Ajiboye et al., 2017; Flesher et al., 2021). Additionally, linear models, due to their small size and simplicity, are easy to interpret and understand, which prevents unexpected behaviors and thus have been the preferred method for applications involving a physical device (Ajiboye et al., 2017; Collinger, Wodlinger, et al., 2013; Flesher et al., 2021).

Non-linear models relax the linear relationship constraint and can instead model potentially arbitrarily complex relationships. This capability aligns with the complex relationship between physiological signals and movement (Abbaspourazad et al., 2023; Sani et al., 2021) and

is further reinforced by the fact that all models with state-of-the-art performance in behavior prediction from physiological signals are mostly non-linear (Azabou et al., 2023; Card et al., 2024; Costello et al., 2023; Deo et al., 2024a; Willett et al., 2021; Willsey et al., 2022; Ye et al., 2023). Since the advent of deep-learning in the 2010s, the non-linear models that have achieved the best performance are based on those architectures: big networks of artificial neurons that can model arbitrarily complex relationships between inputs and outputs. These models have enabled handwriting and speech (Card et al., 2024; Willett et al., 2021), bimanual manipulation (Deo et al., 2024a), control of individual fingers (Costello et al., 2023; Willsey et al., 2022), and even prediction of behavior across subjects and laboratories (Azabou et al., 2023). However, these models' complexity makes researchers regard them as 'black boxes' and has slowed down their adoption in more safety-critical applications (Ajiboye et al., 2017; Flesher et al., 2021).

1.5 Summary Of Thesis

In this thesis, I investigate motor control from three perspectives. First, from a low-level biomechanics perspective, I aim to understand the specific mechanisms that contribute to efficient walking and standing. This involves the impedance of joints involved in those activities, as well as its reliability. Second, from a high-level motor learning perspective, I explore the mechanisms that enable sequential motor learning at a micro-level, trying to understand how individuals acquire and refine motor skills through practice. Finally, I approach motor control from a practical perspective, evaluating how various algorithms can predict motor output from two sources of information: intracortical brain signals and muscle signals. In doing so, I focus on improving the safety and accuracy of these prediction algorithms, which have significant implications for applications in prosthetics and neuroprosthetics.

In Chapter 2, I investigate the test-retest reliability of measuring ankle impedance parameters during standing and walking by producing controlled perturbations at specific times and recording the resulting ankle responses. This study provides insights into the biomechanics of ankle stability, which is key for maintaining balance during walking and standing. Understanding these parameters can inform the design of better assistive devices as well as measure the impact of specific clinical treatments.

In Chapter 3, I address a common problem in the field of motor learning: the lack of programming experience among researchers, which limits their access to tools for testing motor learning hypotheses. To bridge this gap, I present an open-source novel, user-friendly platform that requires no coding skills, allowing researchers to conduct motor learning experiments without the technical barriers traditionally associated with software development. To validate the platform, I replicate a recent study on micro-learning, demonstrating its capability to effectively test motor learning hypotheses and facilitate broader participation in this field.

In Chapter 4, I study how different dimensionality reduction methods compare in their ability to denoise electrophysiology data and improve decoding accuracy. I provide insights into which methods are most effective at compressing and potentially denoising physiological signals. This work has direct applications for improving brain-machine interfaces and other neuroprosthetic technologies.

In Chapter 5, I explore how to combine a principled approach, namely the Kalman Filter, with modern deep learning techniques to produce accurate motor output predictions from intracortical brain data while maintaining algorithm explainability. The aim here is to strike a balance between the predictive power of deep learning and the interpretability of traditional algorithms, ensuring that the resulting models are both effective and explainable. This hybrid

approach can enhance the safety and applicability of brain-controlled prosthetics by making their decision-making processes more transparent.

Finally, in Chapter 6, I discuss how the results presented in this thesis contribute to our understanding of motor control, particularly in relation to prosthetics and neuroprosthetics. I propose future research directions, including the integration of more advanced sensor technologies and the use of adaptive algorithms to enhance user experience. I also share my vision for the future of the field.

Chapter 2 Reliability and Minimal Detectable Change of Stiffness and Other Mechanical Properties of the Ankle Joint in Standing and Walking[1]

2.1 Summary

2.1.1 Background

Ankle joint stiffness and viscosity are fundamental mechanical descriptions that govern the movement of the body and impact an individual's walking ability. Hence, these internal properties of a joint have been increasingly used to evaluate the effects of pathology (e.g., stroke) and in the design and control of robotic and prosthetic devices. However, the reliability of these measurements is currently unclear, which is important for translation to clinical use.

2.1.2 Research Question

Can we reliably measure the mechanical impedance parameters of the ankle while standing and walking?

2.1.3 Methods

Eighteen able-bodied individuals volunteered to be tested on two different days separated by at least 24 hours. Participants received several small random ankle dorsiflexion perturbations while standing and during the stance phase of walking using a custom-designed robotic platform. Three-dimensional motion capture cameras and a 6-component force plate were used to quantify

[1] A version of this chapter "Reliability and minimal detectable change of stiffness and other mechanical properties of the ankle joint in standing and walking" was published in Gait & Posture in February 2024 (Cubillos, Rouse, et al., 2024)

ankle joint motions and torque responses during normal and perturbed conditions. Ankle mechanical impedance was quantified by computing participant-specific ensemble averages of changes in ankle angle and torque due to perturbation and fitting a second-order parametric model consisting of stiffness, viscosity, and inertia. The test-retest reliability of each parameter was assessed using intraclass correlation coefficients (ICCs). We also computed the minimal detectable change (MDC) for each impedance parameter to establish the smallest amount of change that falls outside the measurement error of the instrument.

2.1.4 Results

In standing, the reliability of stiffness, viscosity, and inertia was good to excellent (ICCs=0.67-0.91). During walking, the reliability of stiffness and viscosity was good to excellent (ICCs=0.74-0.84) while that of inertia was fair to good (ICCs=0.47-0.68). The MDC for a single subject ranged from 20%-65% of the measurement mean but was higher (>100%) for inertia during walking.

2.1.5 Significance

Results indicate that dynamic measures of ankle joint impedance were generally reliable and could serve as an adjunct clinical tool for evaluating gait impairments.

2.2 Introduction

The regulation of internal mechanical properties of the ankle joint (i.e., mechanical impedance) is important for the safe and effective ambulation of an individual in the community (Epro et al., 2018; Ho & Bendrups, 2002; S. Kim & Lockhart, 2012). Accordingly, clinicians and scientists have quantified these properties in “healthy” as well as in various patient populations using approaches ranging from simple observation of resistances felt during passive movements to measuring torque responses during controlled perturbations via sophisticated devices (Adjei et

al., 2020; Lee et al., 2014; Nalam et al., 2021; Nalam & Lee, 2017). Collectively, these studies have improved our understanding of the fundamental control and regulation of human movement and locomotion.

From a clinical standpoint, ankle joint impedance has been long used to evaluate impairments after a neurological injury (e.g., spasticity after stroke) and to monitor the efficacy of treatments. Clinicians often quantify joint impedance by qualitatively evaluating the stiffness/resistance observed during passive movements (e.g., Modified Ashworth Scale (Bohannon & Smith, 1987; Meseguer-Henarejos et al., 2018; van Wijck et al., 2001)). A major advantage of these tools is that they are simple and easy to use in the clinic. However, these clinical tools are not objective (Shorter et al., 2021) and provide only a rough estimate of the patient's impairment (Ansari et al., 2008; F. Li et al., 2014; Zurawski et al., 2019). In addition, these qualitative evaluation tools are often performed under static postural conditions and cannot provide insight into the patient's internal properties, which play a role in mediating their observable symptoms. To address these issues, scientists have proposed using joint impedance measurements obtained from measuring the torque responses resulting from an external perturbation (Hunter & Kearney, 1982; Kearney & Hunter, 1982).

Mechanical impedance, defined as the dynamic relationship between perturbation-induced changes in joint angle and the resulting torque response, provides an objective measurement of the internal mechanical properties of a joint. The mechanical impedance of a joint is primarily regulated through the activation and co-activation of the agonist and antagonist muscles and is an important factor that determines an individual's ability to adapt to sudden changes in dynamic tasks, such as during standing or walking. The mechanical impedance can be computed by fitting a second-order system to a joint's torque response curve following a

perturbation (Joshi et al., 2022; Lee et al., 2016). The coefficients of the fitted second-order system are referred to as the mechanical impedance parameters (*i.e.*, inertia, viscosity, and stiffness). These parameters have been shown to differ between "healthy" adults and clinical populations such as stroke (Roy et al., 2011; Shorter et al., 2021) and have been recommended as a surrogate measure to document a patient's progress through rehabilitation. Additionally, an understanding of the normal modulation of joint impedance has provided insight into joint control strategies and has been instrumental in the design of advanced prosthetic devices to emulate human movement (Azocar et al., 2020; Best et al., 2023).

While the mechanical impedance during different phases of walking has been well-characterized and validated thoroughly (Lee et al., 2016), the reliability and sensitivity of these mechanical impedance measurements are currently unknown. This limits the utility of joint mechanical impedance as a progress-tracking tool in the clinic and affects the parameter estimates for prosthetic devices. Further, an understanding of the reliability and minimal detectable change of an instrument is critical before it can be recommended or widely used in the clinic or research. Therefore, the purpose of this study was to evaluate the test-retest reliability and minimum detectable change of the mechanical impedance measurements of the ankle joint during quiet standing and the stance phase of gait in healthy adults. We hypothesized that ankle joint impedance measurements would have good reliability during standing and walking and that the reliability coefficients would be higher during standing than during walking because of the greater inherent control in the experimental parameters, such as foot placement and perturbation timing, that could affect impedance estimates.

2.3 Methods

2.3.1 Participants

Eighteen (9 males, 9 females) young able-bodied individuals participated in this study (age: 22.4 ± 2.8 years; height: 173.3 ± 9.9 cm; mass: 71.2 ± 13.8 kg; 17 right-leg and one left-leg dominant as determined by their preferred leg to kick a ball (Krishnan et al., 2017, 2019)). Participants were tested on their dominant ankle on two different days in two conditions: (1) standing and (2) walking. The study was approved by the University of Michigan Institutional Review Board, and all participants provided written consent before participating.

2.3.2 Perturberator Device

The experiment was conducted using a raised walkway integrated with a previously developed and validated robotic perturbation platform (*i.e.*, the Perturberator) (Arun et al., 1987; Shorter & Rouse, 2018). The Perturberator surface consisted of two portions: a stationary portion (black surface in Figure 2.1A) and a perturbating portion (blue surface in Figure 2.1A). The stationary portion was a steel platform with a rubber mat that was meant to connect the parts of the walkway on either side of the Perturberator. The perturbating portion was fixed to a motorized hinge with its axis located above the surface of the platform so that it roughly aligned with the participant's ankle axis. When activated, the Perturberator generated a small movement in the motorized hinge that would produce a small perturbation about a participant's ankle when they stood/walked on the perturbation portion of the platform (constant velocity ramp; amplitude: 2 deg; duration: 75ms; hold time: 1 sec.; velocity: 45.8 deg/s; max accel: 1800 deg/s^2). These perturbation parameters were chosen to be consistent with previous research and to ensure that the perturbation was large enough to induce a torque response but at the same time did not cause balance or safety issues (Kearney & Hunter, 1982). The velocity and acceleration profiles were chosen to provide a velocity ramp that achieved the 2 degrees within a 100 ms analysis window and complied with the mechanical limits of the motor. Additionally, a six-

component force plate was built into the surface of the perturbing portion of the robot (Type 9260AA, Kistler Inc., Novi, Michigan, USA) that measured the interaction force between the foot and the platform.

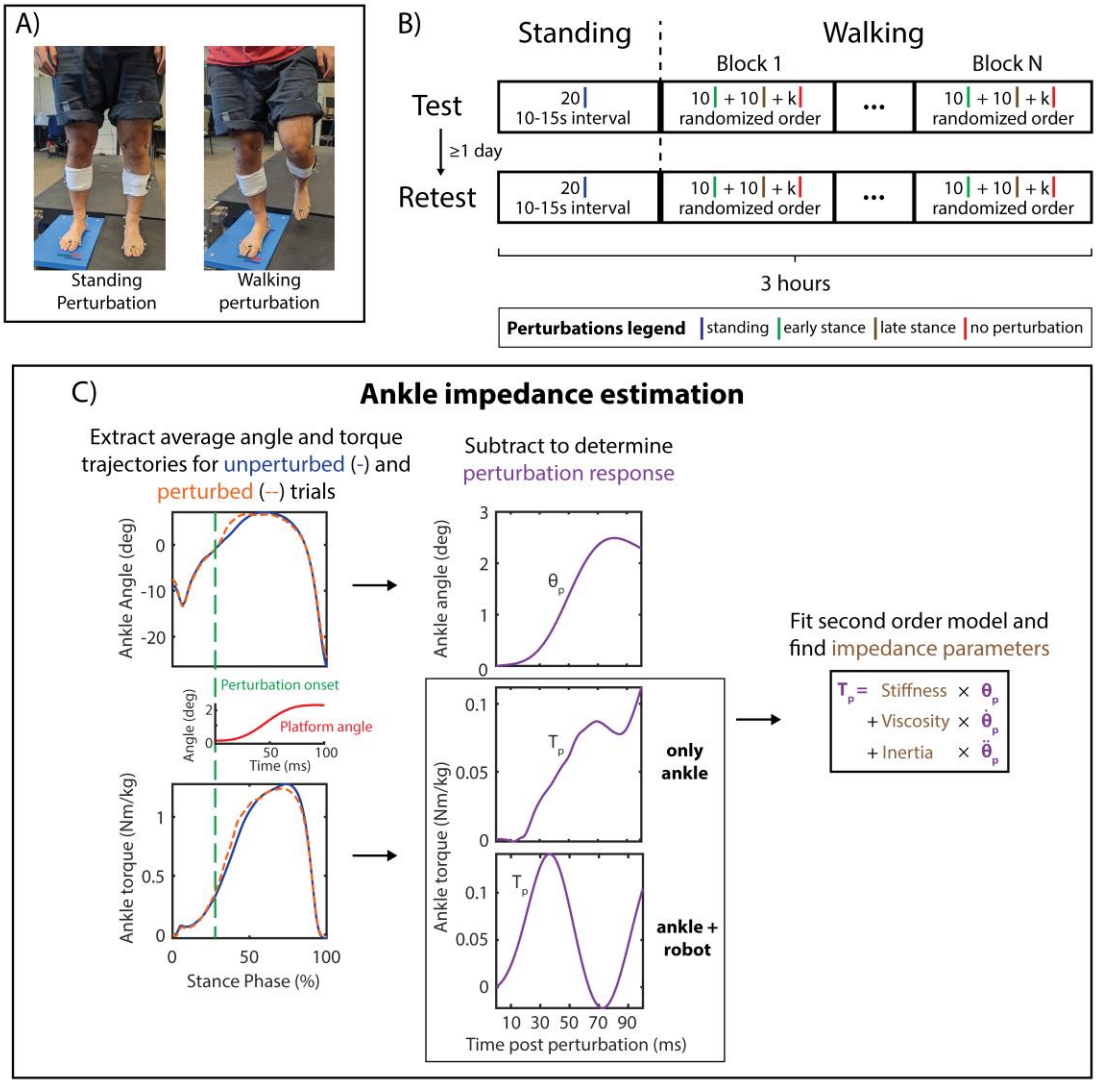


Figure 2.1: **Experimental protocol and ankle impedance estimation method.** (A) Each session was split into a standing block and multiple walking blocks. The left image shows an example of a standing perturbation, while the right one is of a walking perturbation, both for a right-leg dominant participant. The blue platform shown in the picture conveys the perturbation and measures the ground reaction forces. (B) The Test and Retest sessions were replicas of each other and were scheduled with at least one day between them. Each session started with a standing block, where participants received 20 perturbations with a 10 to 15 seconds inter-perturbation delay. During a walking block, 20 perturbations were applied: 10 during the early stance and 10 more during the late stance. For each walking trial, there was a 50% chance of applying a perturbation, which means that not all trials contain a perturbation (represented by k unperturbed trials in the figure). (C) Description of the method used to estimate impedance from measured ankle angle and torque while walking. First, ankle angle and torque (with and without the robot's inertia) are extracted from perturbed and unperturbed trials; the platform angle for a sample perturbation is shown in red (left). Then, the ensemble trajectory of the unperturbed trials is subtracted from each perturbed trial

trajectory to view the perturbation-only response, and a 100ms window around the perturbation is extracted (10ms before perturbation, 90ms after perturbation; middle part of the figure). The average subtracted angle (with the offset removed) is then numerically differentiated (Scheid, 1968) to extract velocity and acceleration. Finally, angle, velocity, and acceleration are used to fit a second-order model to the average subtracted torque response (with the offset removed) and find the impedance parameters (right).

2.3.3 Experimental Protocol

Participants were tested on two different sessions (Test and Retest), each session lasting up to three hours and separated by at least one day (6.2 ± 5.3 days). Each session was separated into a standing condition and a walking condition (Fig. 2.1A). During the standing condition, participants stood motionless on the platform while receiving 20 dorsiflexion perturbations to their dominant ankle that were separated by intervals sampled from a uniform random distribution (lower bound: 10 s, upper bound: 15 s) (Fig. 2.1B). During the walking condition, participants walked on the platform several times unidirectionally so that their dominant foot landed on the Perturberator. Prior to the walking condition, participant's average cadence (steps/second) was determined by having them perform three trials of 10-meter walk test using standardized procedures. This average cadence was used to set the tempo of a metronome. The metronome was used to ensure that the participants maintained a constant stance duration throughout the walking condition in both testing sessions. The participants were then instructed to practice walking to the tempo of the metronome until they felt comfortable matching the rhythm and correctly placing their dominant foot on the perturbing platform (*i.e.*, to land their foot with their ankle axis aligned with the Perturberator's axis). Following practice trials, each participant walked across the platform five times while receiving cues from the metronome to determine the average dominant leg stance duration and early (30%) and late (60%) stance perturbation timings. During the walking session, participants performed six to seven blocks of walking trials. A block continued until 20 perturbations were delivered (10 during early stance,

10 during late stance) to the dominant ankle. Each walking trial had a 25% chance of an early stance perturbation, a 25% chance of a late stance perturbation, and a 50% chance of no perturbation (*i.e.*, an unperturbed trial) to eliminate any anticipatory alteration in the participants' walking behavior. Trials within the block were visually inspected to ensure for correct foot placement and additional trials were performed as needed. Further, during processing, any trial with discrepancy of ≥ 3 cm in the antero-posterior direction was discarded from the analysis. The Retest session was a replica of the Test session, and the self-selected tempo and the stance duration timings from the Test session were maintained for the Retest session.

2.3.4 Impedance Measurements

The force plate angle and ground reaction forces between the participant's foot and the perturbation device were measured (sampling rate of 2000 Hz) with the Perturberator's encoders and a six-component force plate, respectively. The participant's kinematics were measured with 12 motion capture cameras (Miqus M3, Qualisys, Göteborg, Sweden; sampling rate of 200 Hz) tracking 31 reflective markers (4 markers on each shank, six on each foot, two on ankle and knee anatomical landmarks, and three on the robotic platform (Joshi et al., 2022)) (Figure 2.1A). The ankle angle was computed by finding the 3D transformation matrices of the foot and shank relative to their static poses (Arun et al., 1987), decomposing these 3D rotations into Euler angles with one rotation occurring about the ankle axis (*i.e.*, line connecting the malleoli markers of the right foot), and then taking the difference of the two angles about the ankle axis. The ankle torque was computed using the ground reaction forces and center of pressure information to compute the moment about the ankle axis. However, the movement of the platform generates an additional moment about the same axis, and since the force plate is attached to the platform, the force and moment measurements are affected, which in turn affects the torque measurement.

This extra moment can be removed by filtering out the inertia of the force plate (Perreault et al., 1999; Rouse, Hargrove, et al., 2013), but the procedure is highly dependent on the calibration of this filter and the assumption that it remains constant across the entire data collection. Hence, we computed the reliability and minimal detectable change for both cases: "ankle + robot" (no removal of platform's inertia) and "only ankle" (removal of platform's inertia). After computing the torque, ankle kinematics and torque were filtered with a 20 Hz low-pass Butterworth filter (4^{th}-order, zero-phase). For walking, we interpolated the kinematics and kinetics collected during each stance phase to 1000 evenly-spaced points (which allows comparing walking bouts with different stance durations) and then computed each participant's ensemble average.

The mechanical impedance of the ankle during walking was computed by fitting the following second-order parametric model to the ankle's torque response in a 100ms window surrounding the perturbation (10ms before to 90ms after the detected perturbation, which was identified as when the platform angle exceeded 0.02 degrees) using non-negative least-squares system identification:

$$T_p = I_{tot}\ddot{\theta}_p + b_a\dot{\theta}_p + k_a\theta_p \quad (2.1)$$

In equation (2.1), T_p is the ankle torque response to the perturbation—which can either include the torque induced by the movement of the platform ("ankle + robot" case), or not ("only ankle" case)—defined as the difference between the ankle torque response measured during perturbation and the ensemble average of the torque of unperturbed trials (Figure 2.1C). θ_p is the ankle angle response to the perturbation, defined as the difference between the ankle angle measured during perturbation and the ankle angle of the ensemble average of the unperturbed trials; $\dot{\theta}_p$ and $\ddot{\theta}_p$ are the first and second derivatives of θ_p, respectively. I_{tot}, b_a, and k_a are the

mechanical impedance parameters of the ankle, i.e., the total inertia of the foot, the ankle viscosity, and the ankle stiffness, respectively. These parameters were found by using a non-negative least-squares fitting approach (Arun et al., 1987). The impedance for the standing trials was computed similarly, except that the ankle angle and torque responses were measured relative to the starting value right before each perturbation, rather than by subtracting from unperturbed segments.

2.3.5 Statistical Analyses

The test-retest reliability of these impedance parameters was assessed using intraclass correlation coefficients (ICCs) (Bartko, 1966; Cicchetti & Sparrow, 1981; Koo & Li, 2016). ICC analysis was performed using a two-way mixed-effects model for single measurement and absolute agreement (Koo & Li, 2016; Washabaugh et al., 2017), and the resulting ICC values were interpreted as follows: Poor (<0.40), Fair (0.40–0.59), Good (0.60–0.74), and Excellent (0.75–1.00) (Cicchetti & Sparrow, 1981). Additionally, Pearson's correlation coefficients (r) were used to measure the linear strength of association between the Test-Retest measurements. Bland-Altman plots (Martin Bland & Altman, 1986) were used to further examine the agreement of the Test and Retest measurements and to visually inspect any systematic errors between Test-Retest measurements. The individual and group minimal detectable changes (MDC_i and MDC_g, respectively, with 95% confidence) for each impedance parameter were also computed to establish the smallest amount of change that falls outside the measurement error of the instrument and can be deemed a true change for a single participant (MDC_i) or group (MDC_g) (Lu et al., 2008; Vet et al., 2001; Washabaugh et al., 2017). The MDC_i values for each parameter were also represented as a percentage of the measurement mean for easier interpretation. The smaller the MDC_i, the better this method detects changes due to interventions or time. MDC_i and

MDC_g of each impedance parameter were determined by equations (2.2), (2.3), and (2.4) (Beckerman et al., 2001; Campanini & Merlo, 2009; Kesar et al., 2011):

$$\mathrm{SEM} = \mathrm{SD} \cdot \sqrt{1 - \mathrm{ICC}} \quad (2.2)$$

$$\mathrm{MD}_i = SEM \cdot 1.96 \cdot \sqrt{2} \quad (2.3)$$

$$\mathrm{MDC_g} = \frac{\mathrm{MDC_i}}{\sqrt{\mathrm{N}}} \quad (2.4)$$

SEM is the standard error of the measurement, SD is the standard deviation of the parameter measurements across both sessions, ICC is the intraclass correlation coefficient of the parameter, and *N* is the number of participants in a group.

2.4 Results

The reliability for standing was good to excellent for stiffness (ankle + robot ICC: 0.87; only ankle ICC: 0.83), viscosity (ankle + robot ICC: 0.91; only ankle ICC: 0.73), and inertia (ankle + robot ICC: 0.67; only ankle ICC: 0.85) (Tables 2.1 and 2.2; see Supp. Table A.1 through 5 for a breakdown by sex, results normalized by body mass × height, and non-normalized values). For walking, the reliability was also good to excellent for stiffness (ankle + robot ICC: 0.74; only ankle ICC: 0.80) and viscosity (ankle + robot ICC: 0.84; only ankle ICC: 0.78) but only fair to good for inertia (ankle + robot ICC: 0.68; only ankle ICC: 0.47) (Tables 2.1 and 2.2). There were generally strong associations between the impedance parameters obtained on the two different days (r=0.71 to 0.91; Figure 2.1 and Supp. Figure A.1), except for the inertia parameter during walking for the "only ankle" case (r=0.47; Supp. Figure A.1). The Bland-Altman plot showed good agreement between Test and Retest in both cases and across variables for standing and walking: most data points were within the 95% confidence interval, and no clear

trend was observed for those that did not fall inside the interval (Figure 2.2 and Supp. Figure A.1). There were also no systematic errors between the measurements obtained on two days.

Table 2.1: Descriptive statistics for the impedance parameters while standing and walking for the "ankle+robot" case. The last column shows the group MDC as a reference for a hypothetical sample size of N=30.

		Test (Mean ± SD)	**Retest (Mean ± SD)**	**ICC [95% CI]**	**MDC_i**	**MDC_g (30)**
Standing	Stiffness (Nm/rad/kg)	1.35 ± 0.73	1.37 ± 0.82	0.87 (0.69-0.95)	0.75	0.14
	Viscosity (Nms/rad/kg)	0.064 ± 0.018	0.064 ± 0.017	0.91 (0.78-0.97)	0.014	0.003
	Inertia (kgm^2/rad)	0.212 ± 0.011	0.213 ± 0.017	0.67 (0.30-0.86)	0.023	0.004
Walking	Stiffness (Nm/rad/kg)	2.73 ± 1.38	2.96 ± 1.23	0.74 (0.55-0.86)	1.84	0.34
	Viscosity (Nms/rad/kg)	0.049 ± 0.023	0.052 ± 0.026	0.84 (0.71-0.92)	0.027	0.005
	Inertia (kgm^2/rad)	0.209 ± 0.023	0.213 ± 0.029	0.68 (0.47-0.83)	0.041	0.007

Table 2.2: Descriptive statistics for the impedance parameters while standing and walking for the "only ankle" case. The last column shows the group MDC as a reference for a hypothetical sample size of N=30.

		Test (Mean ± SD)	**Retest (Mean ± SD)**	**ICC [95% CI]**	**MDC_i**	**MDC_g (30)**
Standing	Stiffness (Nm/rad/kg)	1.49 ± 0.69	1.50 ± 0.77	0.83 (0.61-0.93)	0.81	0.15
	Viscosity (Nms/rad/kg)	0.038 ± 0.007	0.038 ± 0.006	0.73 (0.41-0.89)	0.009	0.002
	Inertia (m^2/rad)	0.007 ± 0.009	0.007 ± 0.007	0.85 (0.63-0.94)	0.009	0.002
Walking	Stiffness (Nm/rad/kg)	2.51 ± 1.21	2.68 ± 1.03	0.80 (0.65-0.89)	1.38	0.25
	Viscosity (Nms/rad/kg)	0.032 ± 0.009	0.035 ± 0.01	0.78 (0.58-0.89)	0.012	0.002
	Inertia (m^2/rad)	0.008 ± 0.009	0.008 ± 0.009	0.47 (0.17-0.69)	0.019	0.003

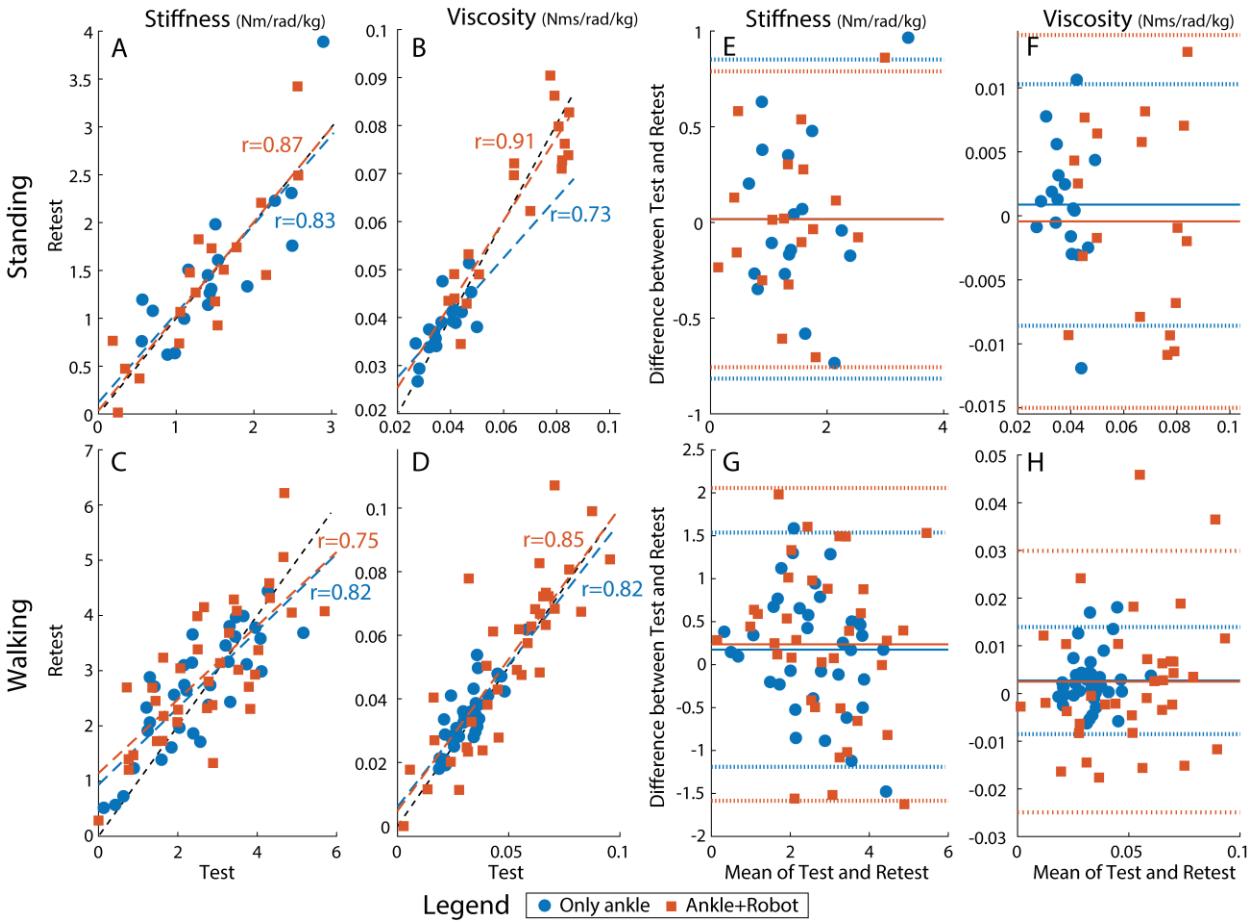


Figure 2.2: **Correlation and Bland-Altman plots for the stiffness and viscosity parameters between sessions for standing and walking**. (A)-(D) show the correlation plots between the Test and Retest sessions for stiffness (first column) and viscosity (second column) during standing (first row) and walking (second row). Blue circles represent values from individual values in the "only ankle" case. Orange squares represent individual values in the "ankle + robot" case. The blue and orange segmented lines represent a linear fit to the data from the "only ankle" and "ankle + robot" cases, respectively, while the black segmented lines show a line of unit slope and zero intercept for reference. The blue and orange text represents the Pearson correlation coefficient between Test and Retest for each corresponding case. (E)-(H) show the Bland-Altman plots for the Test and Retest sessions for stiffness (first column) and viscosity (second column) during standing (first row) and walking (second row). Blue circles and orange squares represent individual values for the "only ankle" and "ankle + robot" cases, respectively. The x-axis on each plot corresponds to the mean between the Test and Retest sessions for each individual value and the y-axis represents the difference between the sessions. The solid blue and orange lines show the mean difference for each case, while the dotted lines represent the 95% confidence interval (1.96 * SD). Stiffness and viscosity showed good correlation and agreement, being close to the dashed black line in the correlation plots in most cases and remaining inside the 95% confidence interval in the Bland-Altman plots. Note for interpretation of the references to color in the figure legend, please refer to the online version of this article.

The stiffness MDC_i ranged from 55 to 65% of the overall stiffness mean for the "ankle + robot" case and from 53 to 54% for the "only ankle" case (Tables 2.1 and 2.2). The viscosity parameter had a smaller MDC_i relative to the measurement mean, amounting from 20 to 55% in the "ankle + robot" case and 24 to 36% in the "only ankle" case. Finally, the inertia MDC_i was

between 10-20% of the measurement means for the “ankle + robot” case but increased to 130-240% on the “only ankle” case.

2.5 Discussion

In this study, we computed the test-retest reliability and minimal detectable change (MDC) of the ankle impedance measured during standing and walking. We found that the stiffness and viscosity parameters had good to excellent reliability (ICC $\geq$ 0.73) for standing and walking and an MDC_i of 20 to 65% of the measurement mean. ICC and MDC for stiffness and viscosity were similar with and without the robot’s inertia removed.

The observed good-to-excellent reliability of stiffness and viscosity shows that these measurements are consistent across test sessions and can be used in the clinic or gait research. The results also have meaningful implications for lower-limb prosthetic research where stiffness and viscosity of the artificial joints during the gait cycle are matched to the experimental results in humans, as this approach could greatly benefit from reliable joint impedance measurements. Additionally, stiffness and viscosity have been shown to differ between stroke and healthy populations (Shorter et al., 2021) and, thus, can be used to complement existing patients’ progress-tracking tools, such as the modified Ashworth scale (MAS) and the modified Tardieu scale (MTS). These clinical assessments are typically measured when the patient is seated or lying down, which limits how improvements in those scales translate to functional improvements (Damiano et al., 2002; Shorter et al., 2021). Our results show that measuring stiffness using perturbations while standing and walking has good to excellent reliability. Thus, complementing the tracking of a patient’s progress with the method tested in this paper may better inform clinicians about what treatments provide functional improvements. However, it is important to note that only healthy participants were tested in this study; further testing with a clinical

population is needed before starting a widespread use of this method as a clinical assessment tool. Additionally, the robot provides a valuable way of measuring changes for research interventions but may need to be simplified further to start using it routinely in the clinic.

The only impedance parameter that showed less than good reliability (ICC < 0.60) was the inertia during walking on the "only ankle" case (Table 2.2 and Supp. Figure A.1), *i.e.*, when the robot's inertia was removed. A potential explanation for this observation could be related to the variations in foot placement during walking and the reduction in the signal-to-noise ratio of the ankle's inertia when the robot's inertia is removed. While walking, participants were instructed to try landing their dominant foot on the platform such that the ankle's axis of rotation matches the robot's axis of rotation. This foot placement is not always perfect (Joshi et al., 2022; Rouse, Hargrove, et al., 2013), which creates some variation in the alignment of the two axes. Since the accelerations induced by the platform are greater than the position or velocity changes, the inertia—the parameter that accompanies the acceleration in the second-order model—is the most affected parameter by these variations. Additionally, the robot's inertia is considerably larger than the ankle's inertia, as can be seen by the difference in mean inertia between the two cases tested (robot+ankle: ~0.210 kgm^2/rad, ankle only: ~0.008 kgm^2/rad). Thus, when removing the robot's inertia (in the "ankle only" case), small variations in ankle inertia between days become big compared to the ankle's inertia, reducing the reliability of the parameter. Further, the use of non-negative constraints resulted in several measurements with '0' values, which also affected the reliability. Despite this limitation, it is important to note that the primary purpose of the inertia estimate is to complete the second-order model and not to have it as a clinical measurement. On the other hand, stiffness and damping measures are not subject to the same limitations—as evidenced by their higher reliability—and are more meaningful as clinical

metrics (Shorter et al., 2021). However, if inertial measurements are of interest, we recommend using the total inertia ("robot + ankle") and a change score from baseline (e.g., after an intervention) to obtain reliable measurements.

The minimal detectable change (MDC) for the impedance measurements was generally higher in walking than in standing (Tables 2.1 and 2.2). This difference is driven mainly by the higher variance of the parameters during walking, likely caused by the differences in foot placement between trials. A lower MDC makes measuring stiffness and viscosity while standing an attractive assessment tool, as the measurements take very little time (around 10 minutes) and do not have the challenges that arise when making the subjects walk and land their foot on a specific location (particularly those with severe walking impairments). Furthermore, in contrast to MAS and MTS (Ansari et al., 2008; Haugh et al., 2006), the measurements during standing are taken in a load-bearing position, which could give greater insight into the symptoms of patients with gait disorders. However, it is important to note that ankle impedance measurements during standing may lack some of the functional insight offered by measuring during gait. Therefore, if a more functional assessment is desired, then measuring during walking may be a better option.

It is important to note that the process of measuring impedance parameters during standing and walking could be affected by the inherent variability in ankle joint position, torque, and step/stride duration. When measuring impedance during standing, standing sway can affect initial absolute ankle angle and torque across perturbations, consequently altering the absolute perturbation-induced ankle torque response curves. However, these variations are small when compared to the perturbation-induced changes in ankle torque (Supp. Figure A.2). Moreover, impedance is assessed using the relative response curves—subtracting the pre-perturbation value

from the response—which show consistent shapes, compared to the unperturbed sway torque (Supp. Figure A.2). When measuring impedance during walking, inter-step variability in foot placement (*i.e.*, aligning the ankle joint axis with the Perturberator axis) and step/stride duration could affect the computed torque and the perturbation timing (Supp. Figure A.3), respectively. We minimized these issues by providing a metronome to control for step duration, providing practice trials and cues to appropriate foot placement via tapes marked on the platform, and also by removing trials in which the ankle axis deviated substantially from the Perturberator axis (≥ 3 cm). Unfortunately, addressing these issues completely is not practical because it would necessitate a perfect estimation of the foot placement and stance duration in real-time. We note that despite these issues, the reliability of the clinically relevant impedance parameters (stiffness and viscosity) was good-to-excellent, which indicates that measuring ankle joint impedance using our current approach is robust to these sources of variability.

Our approach of quantifying ankle joint impedance in standing or walking using system identification techniques remains primarily a lab-based approach, as it requires specialized equipment (robotic platform, motion capture system, force plates) and a multitude of walking trials. This lab-based approach is critical for making precise estimates of ankle impedance and for studying fundamental questions regarding joint impedance, as it provides inherent control and the rigor needed for such processes. From a clinical perspective, such precise estimation may not be necessary as long as the measure can capture the changes dynamically in a functional task with adequate reliability and sensitivity. This could be achieved by reducing the number of trials or performing only standing trials. Moreover, the experiments can be simplified by using a torque-controlled dynamometer in seated or lying posture. However, if a standing or walking posture is preferred due to its clinical relevance (e.g., when studying balance or gait

impairments), the clinical translatability could be improved if less expensive and commonly used clinical tools could be leveraged to study impedance during gait. Along these lines, many large clinics have access to electrogoniometers or affordable motion capture systems that can measure ankle joint motion during walking. Similarly, several low-cost force plates are now available (e.g., 3B scientific) to study kinetics during gait. Thus, estimating the slope of the torque-angle curve (i.e., quasi-stiffness), could be a relatively simple metric that can be used in the clinic as a proxy for joint stiffness. Quasi-stiffness, though distinctly different from joint stiffness when the joint is producing net-positive energy (Rouse, Gregg, et al., 2013), has been widely used in clinical research related to gait abnormalities. However, a more in-depth evaluation of the relationship between the two measures (stiffness and quasi-stiffness) and quantifying whether a change in one metric leads to a change in the other may significantly improve the translatability of our approach.

Note that the current study only tested the reliability of the participants' dominant leg during dorsiflexion perturbation at two phases of the gait cycle. This was intentional, because testing both legs in both directions at multiple phases of the gait cycle would require thousands of trials, which may not be practical without introducing significant confounding effects such as fatigue or boredom. Previous research has shown that internal properties of the joint are similar between legs and perturbation directions (Aird et al., 2012; Rouse et al., 2014) but varies with different phases of gait cycle. Thus, our results should be extrapolated with caution for other conditions that were not tested in this study without further research.

In conclusion, our results indicate that the ankle joint impedance measurements are generally reliable during standing and walking. The reliability of ankle joint stiffness and viscosity was higher than inertia. Similarly, the impedance measures obtained in standing were

generally more reliable than walking. These findings suggest that the ankle joint impedance measures obtained using the procedures described in this study can be a valuable tool for clinicians to measure ankle stiffness and viscosity during standing and walking. Future studies are necessary to determine whether these results generalize to a patient population.

Chapter 3 Breaking the Barriers to Designing Online Experiments: a Novel Open-Source Platform for Supporting Procedural Skill Learning Experiments[2]

3.1 Summary

3.1.1 Background

Motor learning experiments are typically performed in laboratory environments, which can be time-consuming and require dedicated equipment/personnel, thus limiting the ability to gather data from large samples. To address this problem, some researchers have transitioned to unsupervised online experiments, showing advantages in participant recruitment without losing validity. However, most online platforms require coding experience or time-consuming setups to create and run experiments, limiting their usage across the field.

3.1.2 Method

To tackle this issue, an open-source web-based platform was developed (https://experiments.neurro-lab.engin.umich.edu/) to create, run, and manage procedural skill learning experiments without coding or setup requirements. The feasibility of the platform and the comparability of the results between supervised (n=17) and unsupervised (n=24) were tested in 41 naive right-handed participants using an established sequential finger tapping task. The study also tested if a previously reported rapid form of offline consolidation (i.e., microscale

[2] A version of this chapter "Breaking the barriers to designing online experiments: A novel open-source platform for supporting procedural skill learning experiments" was published in Computers in Biology and Medicine in March 2023 (Cubillos et al., 2023)

learning) in procedural skill learning could be replicated with the developed platform and evaluated the extent of interlimb transfer associated with the finger tapping task.

3.1.3 Results

The results indicated that the performance metrics were comparable between the supervised and unsupervised groups (all p's > 0.05). The learning curves, average tapping speeds, and micro-scale learning were similar to previous studies. Training led to significant improvements in mean tapping speed (2.22 ± 1.48 keypresses/s, $p < 0.001$) and a significant interlimb transfer of learning (1.22 ± 1.43 keypresses/s, $p < 0.05$).

3.1.4 Conclusions

The results show that the presented platform may serve as a valuable tool for conducting online procedural skill-learning experiments.

3.2 Introduction

Motor learning is a fundamental and essential part of human behavior, characterized by acquiring new motor skills and enhancing the performance of old ones. Investigations probing the mechanisms that underlie the way we learn and adapt movements (i.e., motor learning studies) have increased dramatically over the past two decades. These studies typically use motor learning tasks that can be broadly categorized into one of the following: motor adaptation (Izawa et al., 2008; Wei & Körding, 2009) or skill acquisition (Higgins, 1991; Krakauer, 2006; Newell, 1991).

Motor learning experiments are typically conducted in person in a laboratory environment to control experimental parameters, but creating this controlled environment comes at a cost (Tsay et al., 2021). In-person experiments require specialized equipment and dedicated personnel to recruit participants and administer experiments. They also impede the ability to

conduct experiments with a large sample size because only one participant can be tested at any given time and some participants may be unable to attend in-lab sessions due to various constraints (e.g., time, cost, transportation). Additionally, in-person experiments typically recruit highly selective participants (e.g., lab members or college students), inducing bias and minimizing the generalizability of the study results (Tsay et al., 2021). Furthermore, lab environments may not always be feasible, such as restrictions imposed by the current global pandemic. Thus, especially given that motor learning studies suffer from issues related to small sample size and publication bias (Lohse et al., 2016), there is an increasing need to look beyond lab experiments.

To address these issues, researchers have turned to online experiments (Anglada-Tort et al., 2022; Bönstrup et al., 2020; Listman et al., 2021; Tsay et al., 2021). Running experiments online makes study recruitment easier for researchers and involvement easier for participants. As a result, investigators can recruit several participants who are more representative of the general population (Paolacci & Chandler, 2014). Online experiments have had great success in the social sciences, with platforms such as PsychoPy (Peirce, 2007), PsyToolkit (Stoet, 2017), Gorilla (Anwyl-Irvine et al., 2020), and lab.js (Henninger et al., 2022) allowing researchers to both create and conduct their studies online. The field of motor learning has also begun conducting online experiments, both by developing new platforms for motor adaptation studies (Tsay et al., 2021) and by using existing platforms for skill learning studies (Bönstrup et al., 2020; Listman et al., 2021). These studies have shown that online experiments are valid and replicate findings from in-person experiments (Bönstrup et al., 2020; Tsay et al., 2021). However, a primary limitation of most current platforms is that they require researchers to code the experiments themselves; thus, users need to have some programming experience. This requirement creates a

barrier to the widespread use of these platforms, ultimately limiting the impact of these online tools on the motor learning research community.

The field of motor learning is broad and uses a diverse set of experimental paradigms (Ranganathan et al., 2021). However, an important requirement for running online experiments successfully is that the experimental paradigms must be simple and should not require equipment or resources beyond those typically available in a participant's home (e.g., computer, mouse, keyboard). One paradigm that fits this need and has been used over the past decade by numerous motor learning researchers is the sequential finger tapping task (Bönstrup et al., 2019, 2020; Korzeczek et al., 2020; Van Der Werf et al., 2009; Witt et al., 2010). In this paradigm, a participant repeatedly taps with their hand a 5-element (or more) sequence (e.g., 41324) of finger movements as quickly and accurately as possible. Performance in this task is typically quantified by the improvements in speed (measured in keypresses/s) and/or accuracy (measured by the number of correct sequences tapped during each trial or block). The sequential finger tapping task has been well-characterized and used extensively in the study of procedural memory formation in the motor learning literature (Bönstrup et al., 2020; Censor et al., 2012). For example, this task has been used to analyze the role of the hippocampus in sequential learning and consolidation of skill (Buch et al., 2021; Diekelmann & Born, 2007), to investigate the mechanisms of offline motor learning at a microscale of seconds (Bönstrup et al., 2019, 2020), and to study the hemispheric lateralization and interlimb transfer of sequence learning (Grafton et al., 2002). Researchers typically rely on their own custom scripts to run these experiments in the lab or online (e.g., using the Amazon Mechanical Turk Platform). However, as mentioned previously, there is a critical need for developing a testing platform that provides flexible options

to perform sequential finger tapping tasks without the need for coding experience to minimize barriers to designing and conducting online experiments.

Therefore, the purpose of this paper was to develop and test a new open-source, web-based platform that allows researchers to design, set up, run, and manage sequential finger-tapping experiments. No coding experience is required, and the experiments can be shared in any crowdsourcing platform to recruit participants online. Notably, the platform requires no specialized equipment or resources, including the need for web hosting. The experimental process is also highly time-efficient because of the ability to quickly create and run online experiments and the availability of pre-processed data immediately after the completion of the experiment. Additionally, a feasibility study was performed to show that 1) the results of in-person and online experiments are similar on the platform and 2) these results are similar to those from a previous seminal paper on a rapid form of offline consolidation in skill learning (Bönstrup et al., 2019). The interlimb transfer (Grafton et al., 2002; Japikse et al., 2003; Lefumat et al., 2015; Parlow & Kinsbourne, 1989; Perez et al., 2007; Wang & Sainburg, 2004) of learning and offline consolidation is also reported, to showcase an example of the new possibilities that this platform gives to interested researchers.

3.3 Methods

3.3.1 *Web-based Experimental Platform*

An open-source web-based platform (https://github.com/lhcubillos/motorlearningapp) was developed to allows researchers in the motor learning field to create and run different types of sequential finger tapping task experiments online without any coding or scripting requirement. The sequential finger tapping task was chosen because this is one of the most commonly used motor learning paradigms that require no specialized equipment for testing purposes, thereby

making it most suitable for online experiments (Bönstrup et al., 2019, 2020; Censor et al., 2012; Friedman & Korman, 2016; Korzeczek et al., 2020; Van Der Werf et al., 2009; Witt et al., 2010). The short sequences used in this task also meant that the learning was focused on how participants improve motor performance, as opposed to other features of learning such as sequence acquisition (Ghilardi et al., 2009).

The platform was developed using the frameworks Django (version 4.0, Django Software Foundation) and Vue (version 3.0, Evan You). Django is a Python-based open-source framework typically used for the back-end development of web applications (i.e., reading and writing to the database, general data processing, etc.). On the other hand, Vue is an open-source JavaScript framework for front-end development that was used to develop the user interfaces (both researcher and participant). The web application is currently hosted on the Google Cloud Platform (GCP) and can be publicly accessed at https://experiments.neurro-lab.engin.umich.edu on any browser. A sample experiment to test the platform is available online at https://experiments.neurro-lab.engin.umich.edu/experiment/5FYA. Due to the platform's open-source nature, researchers with some programming experience can also run the application locally and add any new features they need for their specific studies.

The platform has two user interfaces: the researcher interface (Figure 3.1 top) and the participant interface (Figure 3.1 bottom). The investigators can access the researcher interface after logging in with their username and password created during their profile registration. After logging in, the researcher interface displays the investigator's profile that contains their current studies and options for creating new ones. Within each existing study, the investigator can view the number of responses, download recorded data, publish the study (make it available to participants), or even delete it when necessary. Existing studies are classified as unpublished or

published. A study that is unpublished means that its study parameters are still being decided upon. In this stage, the study can be edited and tested but is unavailable to participants. After the investigator is satisfied with their study, they can publish it. When the study is published, it is available to participants and can no longer be edited by the investigator. Once the investigator is satisfied with the number of participants, or they wish to pause data collection for whatever reason, they can disable the study, making it unavailable to participants. A disabled study can be re-enabled at any time if the investigator wishes to resume data collection.

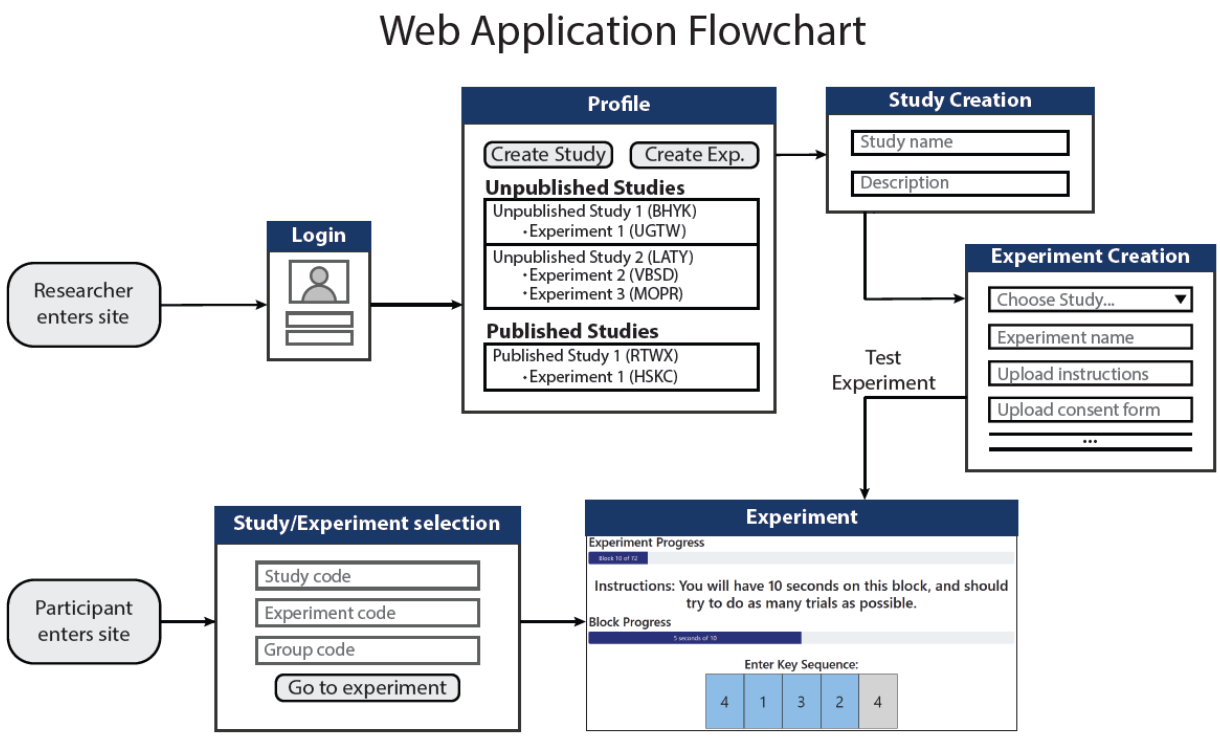


Figure 3.1: **Web application flowchart**. A schematic of the webpage flow as seen by the two types of users: researchers (top) and participants (bottom). Researchers log in to the platform and are then presented with a profile page. There, they can create either a study or an experiment, manage unpublished and published studies, and test experiments (mock examples shown). Participants enter the study, experiment, and group codes given to them by the researchers, and then are redirected to the corresponding experiment.

When creating a new study or editing an unpublished study, the investigator's user interface allows the investigator full control of the study parameters. Within a study, the investigator can select the number of groups that will be included in the study and define the experiments that

each group will do. For example, an investigator could create a simple interlimb transfer study with one group that performs two experiments: one experiment with their right hand and then a second experiment (e.g., a duplicate of the first) with their left. Each experiment is divided into blocks, the number of which is specified by the investigator, and each block is defined by a fixed time interval or a fixed number of key-pressing sequences. Each key pressing sequence is referenced as a "trial" in the platform and all trials in a block have the same sequence. This key sequence can be either manually defined or randomly generated by the program. Through the user interface, the investigator can also control the participant's access to performance feedback during the experiment (both correctness of the most recently pressed key and improvement across blocks) and the rest time (both between blocks and between trials within each block). The platform can also facilitate study execution by allowing the investigator to include instructional videos for the participants and the informed consent document so that participants can view and sign it online.

In the participant interface, the user is first shown a text box, in which they type a study code provided by the investigator. Once the participant submits the code, they are shown the study inclusion/exclusion criteria and the informed consent document, which they are asked to agree to. After consenting to the study, they are shown a short instructional video detailing the experiment instructions, and then redirected to the start of the experiment. While participating in the experiment, the key sequence is presented on the screen and the participant is instructed to begin typing the sequence as instructed (e.g., "type as quickly and as accurately as possible"—these instructions can be customized based on the researcher's need and experimental protocol). Behind each character in the sequence is a grey box. The program offers feedback on the number of keys pressed in the current sequence by changing the color of the grey box behind the current

character in the sequence when a key is pressed. If the investigator enables feedback regarding the correctness of the keypresses, the box will turn blue when the pressed key is correct or orange when incorrect. If this feedback is disabled, the box will turn blue when a key is pressed, regardless of correctness. Once the background color behind a character has been changed, the new color is held until the end of the sequence, after which it resets to the default grey. If the investigator enables feedback on performance for each block, the participant will see a bar graph below the sequence with the number of correct and incorrect trials in the current block. Additionally, regardless of the experiment configurations, the participant will see their progress in the current experiment and their progress in the current block, as well as a short instruction passage to help remind participants what their objective is and what hand they should use for the current block (again, these instructions can be removed from display if needed). When the participants get to the end of the experiment, they are shown a short voluntary survey to fill out. There, the researcher can have the participant fill out some non-identifiable demographic information about themselves, such as age or gender.

Once the investigator is satisfied with the number of responses in a certain experiment, they can proceed to download the participants' data. Three different comma-separated-values (CSV) files can be downloaded, each with different information about the experiment. First, the survey data file provides responses of each of the participants to the survey questions, as well as the time when they started the experiment. Second, the raw data file provides the timestamp and value for each key that the participant pressed, as well as if it was correct in the context of the respective trial. Finally, the processed data file provides the data aggregated in trials, showing correctness and mean/average tapping speed for each trial and participant.

3.3.2 *Feasibility study*

A feasibility study was conducted to compare supervised and unsupervised learning (i.e., participants learning the task in their daily living environment) and to showcase the possibilities of the presented platform to conduct other similar motor learning studies. A total of 43 naïve adult participants (25 males, 18 females, age 30.46 ± 9.8) participated in this feasibility study. All participants were right-hand dominant with no history of significant neurological or orthopedic disorders. The study was reviewed and determined to be Exempt by the University of Michigan Human Subjects Institutional Review Board. All participants agreed to participate in the study with an online consent form before starting the experiment.

Participants were separated into two groups: a “supervised” group and an “unsupervised” group. Participants in the “supervised” group (18 participants, 10 males, 8 females, age 30.1 ± 10.9 years) performed the experiment in a controlled, laboratory environment while under the supervision of a study team member. Participants in the “unsupervised” group (25 participants, 15 males, 10 females, age 30.8 ± 9.2 years) performed the experiment outside of the laboratory (typically in their homes) via the internet, without supervision. Participants in the supervised group were mostly recruited from the university pool and had relatively similar background (college students), whereas participants in the unsupervised group were more diverse and were mostly outside the university pool to ensure that they represented a more heterogenous sample that are typically observed in the “wild” (i.e., ecologically valid settings). Except for the differing levels of supervision, all experimental protocols were identical between the groups.

The experiment consisted of two parts with one minute of rest between them. In the first part of the experiment, participants were instructed to accurately type the sequence '41324' as fast as possible with their non-dominant (left) hand [Figure 3.2A left]. While a shorter sequence like this may seem simple, it is important to note that the primary purpose of this study was to

evaluate how the speed at which the participants accurately typed the sequence improved with practice, not in accuracy itself. Moreover, using a simpler sequence allowed participants to focus on motor performance improvements rather than other features of learning, such as sequence acquisition (Ghilardi et al., 2009). More importantly, this first part of the experiment was a replication of a previous study (Bönstrup et al., 2019) to directly compare the proposed platform's results with previous research without the confounding factor of a new experimental design. Each participant received 36 blocks of training with each block consisting of 10 seconds of practice and 10 seconds of rest. During practice, the participants were instructed to accurately type as many sequences as possible. Whenever a key was pressed, the platform informed the participant of the total number of keys pressed in each sequence, but not whether they were correct or not. Participants were instructed to look at their computer screens (not at their hands) throughout the experiment. In the second part of the experiment, participants typed the sequence '70897' as fast and as accurately as possible with their right hand [Figure 3.2A right]. Other than the sequence and the hand used for training, the second part of the experiment was identical to the first. The addition of the second part of the experiment enabled using the platform to examine the interlimb transfer of learning from the first part of the experiment. The sequence was chosen to be the mirrored version of the first sequence, as previous research has shown that interlimb transfer is present when the second sequence is mirrored from the first (Grafton et al., 2002). Note that the objective of the second part of the experiment was to demonstrate the feasibility of the platform's ability to quickly and easily conduct other forms of motor learning studies and answer a variety of motor learning research questions.

3.3.3 Data Analysis

All outcome metrics were computed from each participant's tapping speed in a trial. A description of all these variables, along with the mathematical notations to compute them, can be found in 6.3Appendix B:. A trial's tapping speed was found using Equation (3.1) (Bönstrup et al., 2019), and it quantified the rate at which the participant was pressing keys:

$$\text{Trial tapping speed} = \frac{1}{\frac{1}{4}\sum_{i=1}^{4} \text{interval}_i} \quad (3.1)$$

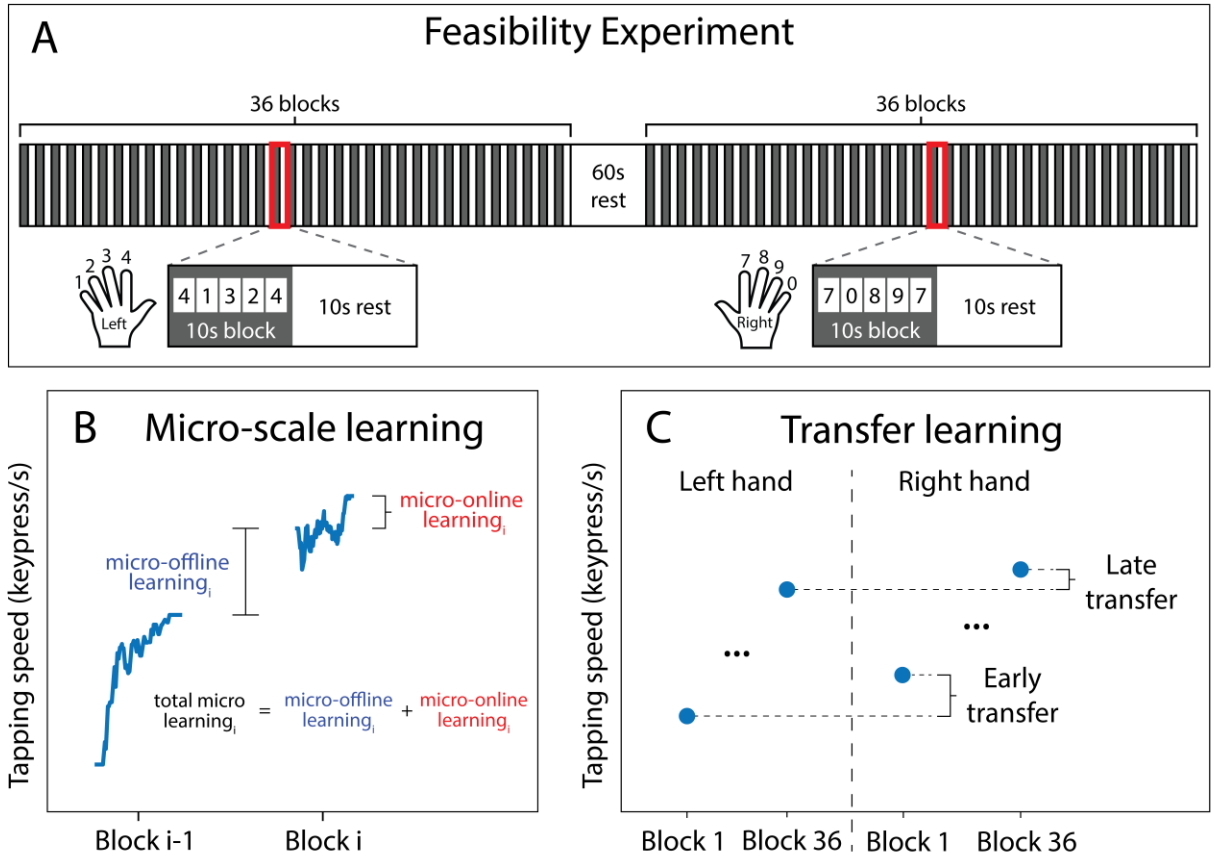


Figure 3.2: **Schematic of experimental protocol and data processing**. (A) A schematic of the experimental protocol. Participants practiced a finger tapping task where they first learned to type the sequence '41324' as fast and as accurately as possible with their non-dominant, left hand. After a one-minute rest following the completion of the task with their left hand, they then learned to type the sequence '70897' as fast and as accurately as possible with their right hand. Participants received 36 blocks of training for each hand with each block consisting of 10 seconds of practice and 10 seconds of rest. (B) A schematic for the computation of micro-scale learning (micro-offline and micro-online learning). Micro-offline learning quantified the learning that occurred during short periods of rest, and micro-online learning quantified the learning that occurred during short periods of practice. For block i, micro-offline learning is defined as the difference in tapping speed (keypresses/second) between the first trial of block i and the last trial of block i-1. Micro-online learning is defined as the difference between the last and first trials of block i. Total micro-learning at that block is the sum of both quantities. (C) A schematic of the interlimb transfer of motor skill learning metrics. Early transfer is calculated as the difference in the mean trial tapping speed between the

first block of the right and left hand. Late transfer is calculated as the difference in mean trial tapping speed between the last block of the right and left hand.

Where interval_i represents the time interval between keypresses i and $i + 1$ in a given trial. Tapping speed was only calculated when the trial was correct (i.e., all keys of the sequence were pressed in the correct order and within the time frame) (Bönstrup et al., 2019, 2020).

The tapping speed was used to compute the following learning metrics: early micro-online learning, early micro-offline learning, early total micro-learning, and total learning across all trials. These metrics were computed for both the right and left hands, generating eight learning metrics. The first 11 blocks of training for the hand under consideration were used for studying early learning (Bönstrup et al., 2019). Micro-offline learning was defined as the difference in tapping speed (keypresses/second) between the first trial of a practice block and the last trial of the previous block (Figure 3.2B), (Bönstrup et al., 2019) and quantified learning that occurred during short periods of rest. Early micro-offline learning was computed as the sum of the micro-offline learning for the first 11 blocks. Micro-online learning was defined as the difference in tapping speed between the last and the first trial of a practice block (Figure 3.2B), (Bönstrup et al., 2019) and quantified learning that occurred during short periods of practice. Early micro-online learning was computed as the sum of the micro-online learning for the first 11 blocks. Total micro-learning was defined as the sum of the micro-offline and micro-online learning and quantified the learning between blocks (i.e., from the beginning of one block to the beginning of the next block). Early total micro-learning was calculated as the sum of total micro-learning in the first 11 blocks and represented all changes in tapping speed during early learning (Bönstrup et al., 2019). Total learning over all trials was calculated as the difference between the mean trial tapping speed of the first and the last block of practice.

To quantify if performance improvements transferred from one limb to the other (i.e., interlimb transfer), two transfer metrics were calculated: early transfer and late transfer (Figure 3.2C). Early transfer was calculated as the difference between the mean trial tapping speed of the left and right hands on the first block of practice. Late transfer was calculated as the difference between the mean trial tapping speed of the left and right hands on the last block of practice.

3.3.4 *Data Management and Analyses*

Data from two participants were excluded from the analysis. One participant from the unsupervised group was excluded because they did not understand the experimental instructions and performed only one trial per block. A participant from the supervised group was excluded because they had a very high mean tapping speed and was identified as an outlier based on Tukey's fences (Tukey, 1977).

An inclusive statistical approach that involved classical (p-value) and Bayesian analyses (Bayes Factor) was used when comparing the results of supervised and unsupervised learning. Learning and interlimb transfer data were compared between the supervised and unsupervised groups using two-tailed two-sample t-tests and two-tailed Bayesian two-sample t-tests. The Bayes factor (BF_{01}) for the likelihood of the null hypothesis (groups are equal) over the alternative hypothesis (groups are different) was used to interpret the results from two-tailed Bayesian two-sample t-tests. P-values were adjusted for multiple comparisons using the Benjamini-Hochberg procedure (BHP) to control for false discovery rate (FDR) (Benjamini & Hochberg, 1995; Glickman et al., 2014; Korthauer et al., 2019). If no significant differences were observed between the groups across all metrics, the data from all participants were collapsed for each of these metrics and a two-tailed one-sample t-test was run to test for practice effects on learning and interlimb transfer. Again, the Benjamini-Hochberg procedure was used to control

for FDR due to multiple comparisons. A significance level of $\alpha = 0.05$ was used for all statistical analyses.

Additionally, a linear discriminant analysis (LDA) model was trained to classify the group for each participant, using the ten learning and interlimb transfer metrics described above as features of the LDA model. The output from the LDA classifier was compared to that from a random classifier (a random classifier simply 'guesses' between the two groups) to determine whether the features contained enough information to discriminate between the two groups. The comparison between the two classifiers was performed using Dietterich's 5x2-fold cross-validation (CV) test (i.e., 5 iterations of twofold cross-validation), which is more powerful and less prone to type I errors than paired t-tests when comparing classifiers (Dietterich, 1998). This process was iterated 1000 times to get a distribution of the accuracy of the classifiers and the p-values. Note that if the null hypothesis is true (no difference in accuracy between LDA and random classifiers), then the p-values should follow a uniform distribution.

3.4 Results

All the data for the supervised (code = 5P6U) and unsupervised (code = C6XN) groups are available here: https://osf.io/k43qw/. Participants in the supervised group had similar learning curves across both hands to those in the unsupervised group (Figures 3.3A and 3.3B). Both groups improved in tapping speeds in the early trials and then tended to stabilize towards the end. The initial tapping speed was generally higher on the right hand when compared with the left hand in both groups. No significant difference between the groups in any of the ten metrics described above were found (all BHP adjusted p-values ≥ 0.65, Table 3.1 and Supp. Figure A.1 and 2). The observed effect sizes were also small, and the BF_{01} values were generally favoring the null hypothesis (Table 3.1). The two classifiers (LDA and random) did not show a significant

difference (52 ± 5% vs. 50 ± 3% (mean ± S.D.); mean p=0.55) when predicting the group of participants (Figure 3.4). The p-values from the Dietterich's test followed a uniform distribution with a mean p = 0.55 with p-values exceeding the significance level ($\alpha = 0.05$) for 97.2% of the time.

Table 3.1**:** Descriptive data for the ten learning metrics in the supervised and unsupervised groups.

Variable	Unsupervised (Mean ± S.D.)	Supervised (Mean ± S.D.)	Effect Size Cohen's d	Bayes Factor (BF_{01})
Left early total micro-learning	1.21 ± 1.70	2.08 ± 1.24	−0.602	0.80
Left early micro-offline learning	4.81 ± 4.89	3.75 ± 4.27	0.234	2.60
Left early micro-online learning	−4.10 ± 5.36	−2.10 ± 5.09	−0.384	1.81
Right early total micro-learning	1.12 ± 1.19	1.41 ± 1.88	−0.174	2.87
Right early micro-offline learning	2.75 ± 5.13	3.31 ± 5.58	−0.104	3.10
Right early micro-online learning	−2.08 ± 6.22	−2.33 ± 6.80	0.038	3.21
Total left learning	2.20 ± 1.81	2.24 ± 1.23	−0.022	3.23
Total right learning	1.02 ± 1.13	1.23 ± 1.54	−0.146	2.97
Early transfer	1.23 ± 1.57	1.21 ± 1.37	0.013	3.23
Late transfer	0.05 ± 0.81	0.20 ± 1.01	−0.162	2.91

S.D. = standard deviation. Note that two-sample t-tests with false discovery rate (FDR) corrections for multiple comparisons using the Benjamini-Hochberg procedure (BHP) indicated that none of the variables were significantly different between the groups (all BHP adjusted p-values > 0.65).

After collapsing data from both groups, all metrics except *late transfer* were significant (Table 3.2). On the left hand, micro-offline learning was positive for all early blocks of training, while micro-online learning was positive only in the first block and near-zero or negative for the remainder of early training (Figure 3.5A). The decomposition of early learning into micro-offline and micro-online learning showed that early learning was initially driven by positive gains in

micro-offline learning but eventually plateaued as gains in micro-offline learning were offset by a loss in micro-online learning (Figures 3.5C and 3.6 left). On the right hand, early learning plateaued earlier, with most learning appearing on the first two blocks (Figure 3.5 B and D, and 3.6 right). The micro-offline and micro-online learning followed a similar pattern as the left hand, except that the extent of micro-offline gains was smaller.

Table 3.2**:** Descriptive data for the ten learning metrics collapsed between groups

Variable	All participants (Mean ± S.D.)	Effect Size Cohen's d
Left early total micro-learning	$1.72 \pm 1.49^{\dagger}$	1.155
Left early micro-offline learning	$4.19 \pm 4.51^{\dagger}$	0.930
Left early micro-online learning	$-2.93 \pm 5.23^{\dagger}$	−0.560
Right early total micro-learning	$1.29 \pm 1.62^{\dagger}$	0.797
Right early micro-offline learning	$3.08 \pm 5.34^{\dagger}$	0.577
Right early micro-online learning	$-2.23 \pm 6.49^{\dagger}$	−0.343
Total left learning	$2.22 \pm 1.48^{\dagger}$	1.503
Total right learning	$1.14 \pm 1.37^{\#}$	0.831
Early transfer	$1.22 \pm 1.43^{*}$	0.849
Late transfer	0.13 ± 0.93	0.146

S.D. = standard deviation. Note that one-sample t-tests with false discovery rate (FDR) corrections for multiple comparisons using the Benjamini-Hochberg procedure (BHP) indicated that all of the variables, except late transfer, were significantly different from 0. Daggers (†) denote $p<0.001$, hashes (#) denote $p<0.01$, and asterisks (*) denote $p<0.05$.

Early transfer (*i.e.,* the difference between tapping speed in the left and right hands in the first block of training) was significantly different from zero (1.22 ± 1.43 keypresses/s, Mean ± S.D., BHP adjusted $p = 0.038$), but late transfer (*i.e.,* the difference between tapping speed in the left and right hands in the last block of training) was not significantly different from zero (0.13 ± 0.93 keypresses/s, Mean ± S.D., BHP adjusted $p = 0.356$).

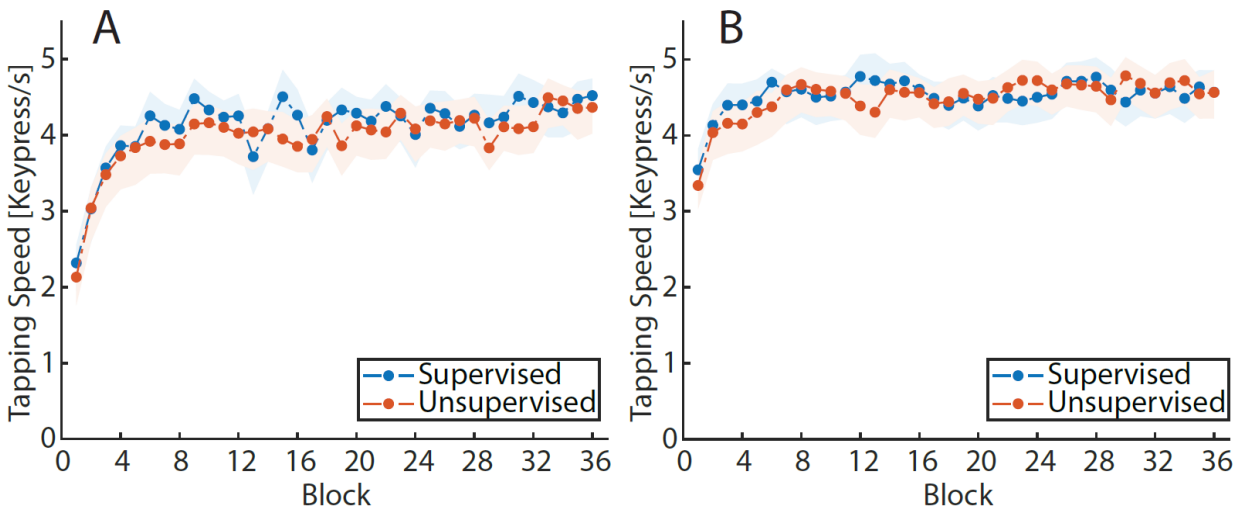


Figure 3.3: **Tapping speed performance**. A scatterplot showing the mean tapping speed (keypresses/second) of the supervised and unsupervised groups in each block of practice for the (A) left and (B) right hand of the participants. The shaded region indicates the standard error of the mean (SEM). Note that the performance on the finger tapping task was similar between the supervised and unsupervised groups for both hands. Also, note that the tapping speed on the initial block of the right (i.e., transfer) hand was higher than the left (i.e., training) hand, indicating the presence of interlimb transfer of procedural skill learning.

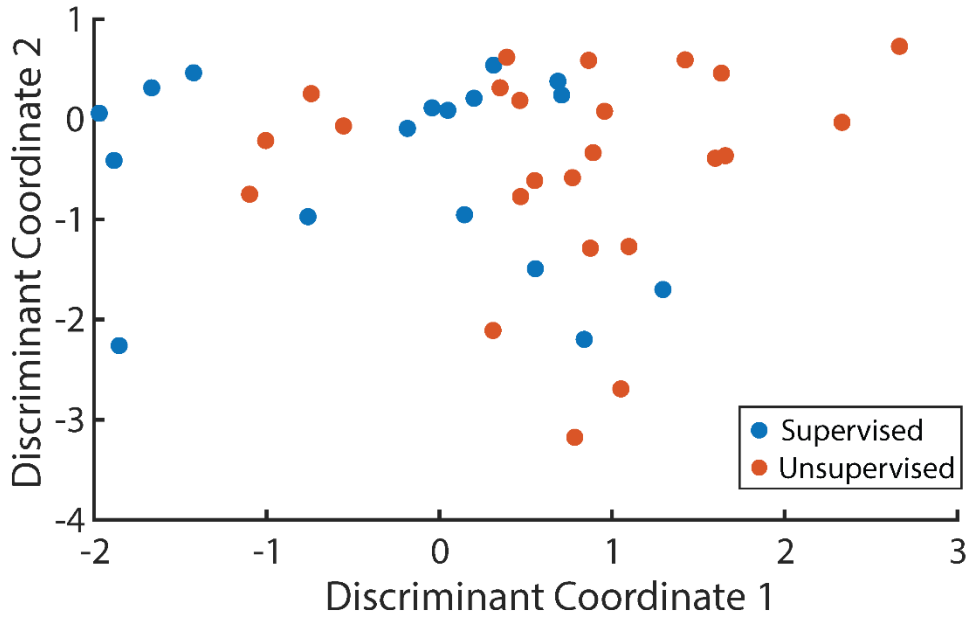


Figure 3.4: **Classification using Linear Discriminant Analysis (LDA) model:** Scatterplot showing the projection of the ten features described in Table 3.1 and Table 3.2 to the first two discriminant coordinates of a linear discriminant analysis (LDA) model. Each marker represents a single participant, and the color represents the group they belong to. No clear separation can be observed between the two groups, and no significant difference in classification accuracy was found between the LDA classifier and a random classifier (mean $p = 0.55$). Note that for visualization purposes only the first two discriminant coordinates have been plotted.

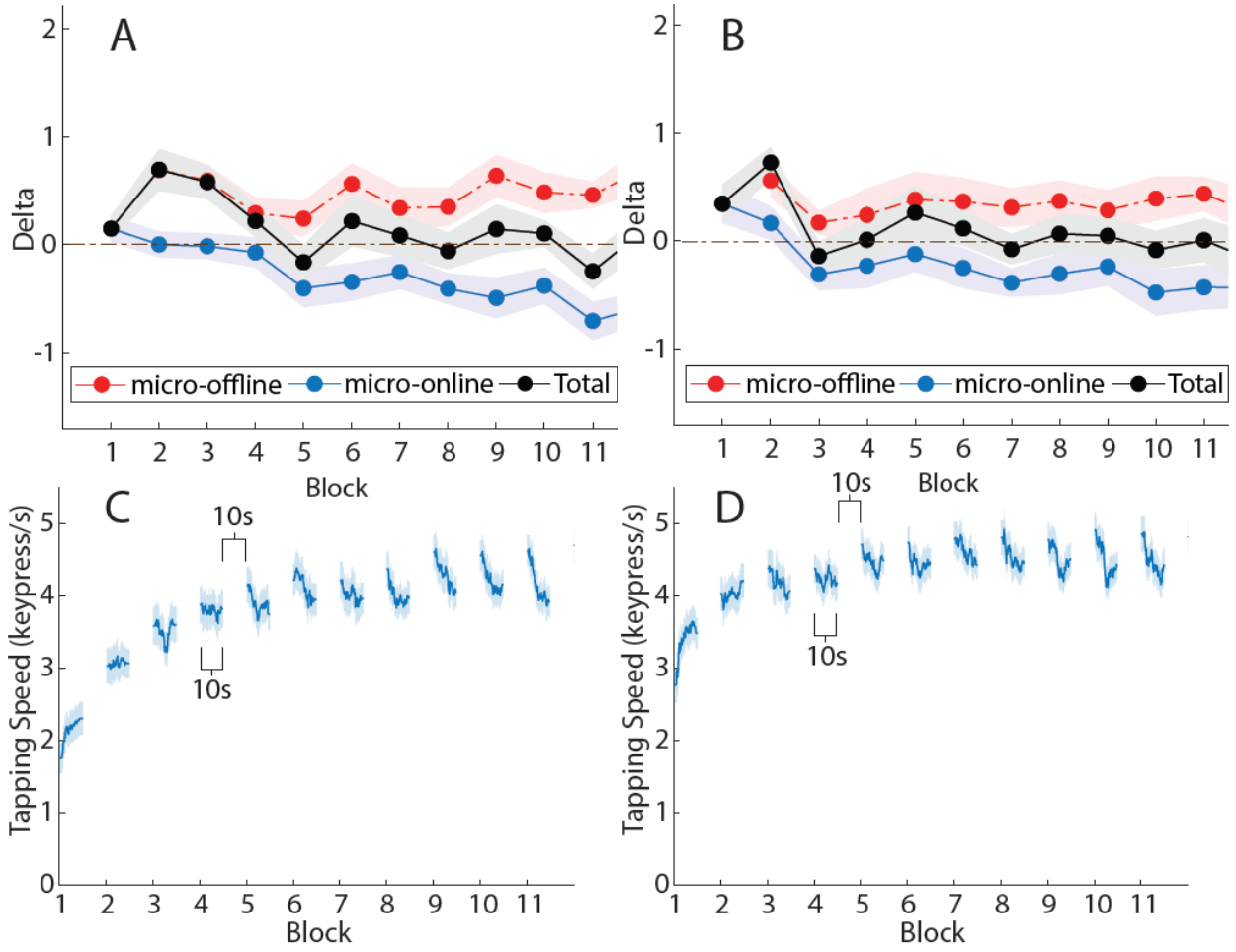


Figure 3.5: **Micro-scale learning**. A scatterplot showing the mean micro-scale (micro-offline, micro-online, and total) early learning (first 11 blocks) for the **(A)** left and the **(B)** right hand across all participants. Micro-offline learning was defined as the difference in tapping speed (keypresses/second) between the first trial of a practice block and the last trial of the previous block and quantified learning that occurred during short periods of rest. Early micro-offline learning was computed as the sum of the micro-offline learning for the first 11 blocks. Micro-online learning was defined as the difference in tapping speed between the last and the first trial of a practice block and quantified learning that occurred during short periods of practice. Early micro-online learning was computed as the sum of the micro-online learning for the first 11 blocks. Total micro-learning was defined as the sum of the micro-offline and micro-online learning and quantified the learning between blocks. Early total micro-learning was calculated as the sum of total micro-learning in the first 11 blocks and represented all changes in tapping speed during early learning. The shaded region represents the standard error of the mean (SEM). The horizontal dashed line shows the zero value. Note that micro-online learning was negative for most of the blocks, indicating that participants had higher tapping speeds at the beginning than at the end of each block. Traces showing the mean tapping speed for the **(C)** left and **(D)** right hand across all participants for the ten seconds in each block during early learning. The shaded region represents the SEM. Note that this plot was created only for visualization purposes and the details for creating this plot are provided in Appendix B:.

3.5 Discussion

This paper introduces a novel open-source web-based application that allows researchers with no coding experience to create and conduct their own motor sequence learning experiments

from anywhere in the world. The feasibility of the application was also established by showing that (1) the results from unsupervised online experiments were comparable to those of supervised in-person experiments and (2) the results were qualitatively in close correspondence with previous studies (Bönstrup et al., 2019, 2020). Finally, the flexibility of the platform was demonstrated by performing an interlimb transfer study to quantify the extent of motor skill transfer between the left and right hand. These findings demonstrate that the presented web-based platform could serve as a viable tool to conduct online procedural skill learning experiments that involve sequential finger tapping.

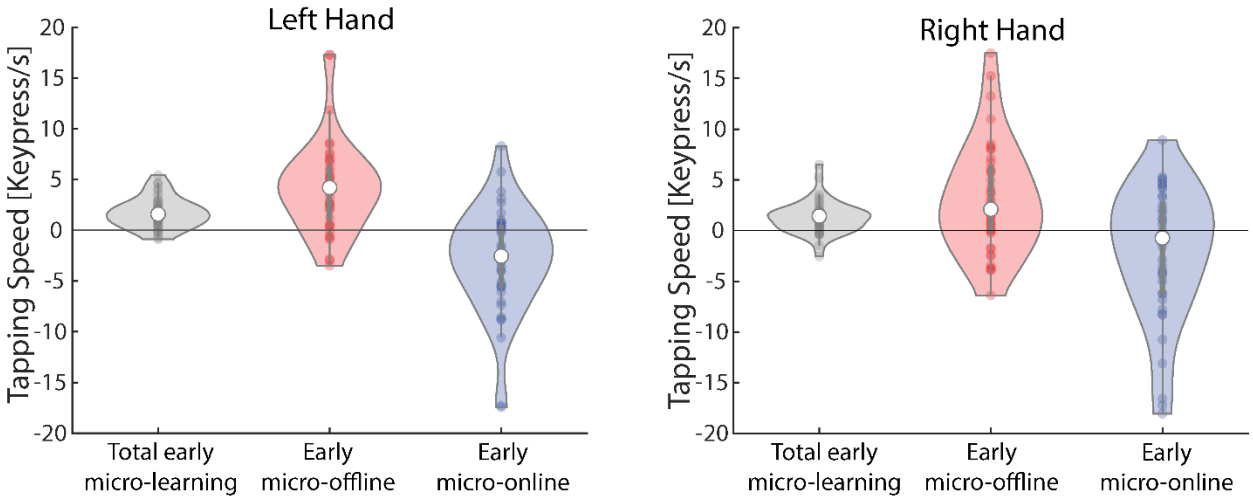


Figure 3.6**: Distribution of micro-scale learning:** Violin plots showing the distributions of early micro-offline learning, early micro-online learning, and early total micro-learning for the **(left)** left and **(right)** right hand across all participants. Early micro-learning (micro-offline, micro-online, and total) was computed as the sum of the values observed in the first 11 blocks. Note that the white circles indicate the median values, the filled circles indicate individual data points, and the width of the shaded region represents the approximate frequency of data points in each region.

Previous studies have evaluated the feasibility of conducting online behavioral experiments ranging from reaction time measurements (Semmelmann & Weigelt, 2017) to motor learning studies involving sequential finger tapping task (Bönstrup et al., 2020) and visuomotor adaptations (Tsay et al., 2021). These studies have found that online experiments yield similar

results to those performed in-person in the lab and can be a suitable alternative to laboratory experiments (Bönstrup et al., 2020; Casler et al., 2013). The results presented here are consistent with these studies, as no meaningful differences were found in any of the measured parameters between the supervised and the unsupervised groups despite differences in background and ethnicity. This finding and those of others indicate that a highly structured laboratory environment may not always be necessary to conduct certain types of motor learning experiments, particularly when no special instrumentation other than a personal computer is needed for the experiment. Furthermore, conducting experiments online is helpful for researchers, as it allows them to collect data more efficiently from a broader subject pool that would otherwise be impossible due to financial or geographical constraints (Paolacci & Chandler, 2014). Recruiting from previously inaccessible subject pools makes study results more generalizable. Running experiments online also usually decreases the data collection time (Bönstrup et al., 2020), as subjects can participate simultaneously and at any time, without laboratory space or work hours constraints.

Participants showed a steep increase in tapping speed in the early blocks, followed by a rapid stabilization in the later blocks. When decomposing the early learning into micro-online and micro-offline learning, the results showed that the initial improvements mostly happened at rest rather than during practice. This finding is consistent with Bonstrup et al. (Bönstrup et al., 2019), who reported a rapid offline consolidation that contributed to early skill learning. While the exact mechanism for this phenomenon is unclear, some have argued that the emergence of offline improvements could be due to the loss of inhibitory drive or dissipation of fatigue during the short rest intervals (Robertson, 2019). Further studies are necessary to understand the mechanisms underlying these results.

The mean tapping speed on the initial block of the right hand was significantly higher than the left hand, indicating a significant interlimb transfer of motor skill learning. Other studies have shown similar inter-limb transfer in both adaptation (Lefumat et al., 2015; Wang & Sainburg, 2004) and skill learning (Grafton et al., 2002; Japikse et al., 2003; Parlow & Kinsbourne, 1989; Perez et al., 2007) tasks. Previous studies on sequential motor learning tasks have shown that inter-limb transfer is present when the sequence for the second hand (i.e., transfer hand) is either the same or a mirrored sequence of the first hand (i.e., training hand) (Grafton et al., 2002). Thus, the observed interlimb transfer in this study may be due to the use of a mirrored sequence between the non-dominant and dominant hand, as opposed to some other random sequence. However, because of the study design, it is unclear whether the increase in tapping speed on the right hand was due to interlimb transfer or due to the advantage of the dominant hand. It is to be noted that no pre-test evaluations were performed to establish the baseline performance of the dominant, right hand because the use of pre-tests in motor learning has been criticized for yielding unreliable scores and providing practice of the task, thereby confounding the results of the intervention (Ranganathan et al., 2022; Schmidt et al., 2018). However, when looking at the learning curves (Figure 3.3), it appears that this observation was primarily due to the transfer of the learned skill from the left hand, as this was only observed during the initial blocks of training. Further research may be needed to determine whether the transfer of learning occurs to the same extent when performing the task first with the dominant and then with the non-dominant hand, albeit, recognizing that this study design may also have limitations due to the presence of asymmetric transfer of learning between hands/hemispheres (Lavrysen et al., 2003; Wang & Sainburg, 2004).

Researchers interested in studying sequential motor learning with a finger tapping task will find the introduced software valuable because of its unique advantages over previously developed platforms. Specifically, the proposed platform has four innovative features. First, the platform requires no coding experience to create and manage sequential finger-tapping experiments online. Second, it is low cost and requires no specialized equipment or resources, including the need for web hosting. Third, it makes the experimental process highly time-efficient, as the set-up process is made easier by pre-hosting the platform on the google cloud platform and the data management process is made easier by pre-processing the data to extract key variables. Finally, the platform is highly flexible to run other forms of sequential finger tapping tasks (e.g., interlimb transfer) without the need for writing new scripts or codes to make this feasible.

Other software, such as PsychoPy (Peirce, 2007), PsyToolkit (Stoet, 2017), Gorilla (Anwyl-Irvine et al., 2020), and lab.js (Henninger et al., 2022), provide researchers with a common platform to create and/or distribute experiments. The four tools have great flexibility in the types of experiments that can be created and provide a way to either directly manage experiments or upload them to other management software, such as Pavlovia (https://pavlovia.org/). However, they have some limitations. PsychoPy requires the researchers to manually set up the necessary tools, which can be a challenge both in experiment creation and distribution. PsyToolkit, on the other hand, is a web-based platform; and thus, does not require any setup from the investigator. However, it requires learning the PsyToolkit scripting code to configure them, which could be a big challenge for most researchers. Gorilla overcomes the limitations of the previous tools, but its cost structure and closed-source nature could be a barrier and prevent the community from being able to actively participate in the development and

prioritization of new bug fixes and features. Finally, lab.js is a recent open-source tool with great flexibility to create different types of motor learning experiments. However, it does not natively allow researchers to conduct online experiments and track participants' responses online, which makes it rely on other tools for that purpose. Furthermore, although finger tapping experiments can be created using their software, it requires researchers to have an intimate knowledge of the platform's documentation, experimental set-up process, and some amount of programming experience.

The software introduced here is an open-source web-based application that requires no setup and allows researchers to create sequential task motor learning experiments with a form-like graphical interface. Although it does not allow the creation of any type of motor learning or behavioral experiment, it greatly simplifies experiments involving finger tapping tasks by allowing the user to choose from an array of options for each possible configuration. Thus, it can facilitate online behavioral experiments for multiple research groups interested in procedural skill learning through sequential finger tapping tasks (Balas et al., 2007; Bönstrup et al., 2019, 2020; Friedman & Korman, 2016; Korzeczek et al., 2020; Van Der Werf et al., 2009; Witt et al., 2010). The platform also simplifies the process of conducting experiments by providing a way of viewing the number of responses in real-time and allowing the researchers to enable or disable each experiment at any given time when satisfied with the number of responses.

The use of the web-based platform could potentially minimize the variation in tasks across studies, thereby reducing task fragmentation across motor learning studies. Scientists have argued that a high level of task fragmentation poses significant theoretical and methodological barriers to advancing the field (Ranganathan et al., 2021). The presented software addresses this issue by offering researchers a common task in the same platform to design and conduct their

motor learning studies. This also allows for easier replication of motor learning studies. To ease this process further, a new feature will be eventually added to the platform that will allow users to export their experiment configurations as a file to be shared. Then, other users will be able to import that file and create a replica of the experiment created by the original researchers.

A limitation of the platform is that it can currently be used only to perform experiments involving sequential finger tapping tasks, making it less flexible than other similar coding-based platforms (Anwyl-Irvine et al., 2020; Henninger et al., 2022; Peirce, 2007; Stoet, 2017). It is to be noted, however, that there is a tradeoff between flexibility and usability. While other platforms are flexible to perform several cognitive and behavioral experiments, they only cater to the needs of researchers with programming/coding experience. Thus, the focus here was on the usability issue, where typical motor learning researchers (e.g., kinesiologists and rehabilitation scientists) find it difficult to use platforms that require coding experience. However, given the infrastructure developed for the platform, it would be relatively simple to expand the software to include other commonly used motor learning tasks with similar technical requirements. For instance, serial reaction-time tasks (Chambaron et al., 2008; Robertson, 2007) could be easily integrated into the platform. In these tasks, stimuli must be shown at appropriate times, and the timing of the participant's responses must be recorded, both technical capabilities that the web application already possesses. Thus, extending the platform to those tasks would be a natural evolution. Furthermore, the platform is being released as open-source with a GNU GPL v3.0 license, allowing researchers to actively participate in the discussion and development of new features relating to online motor learning experiments. Another limitation is that there are no ready-made templates or a feature to export and import the experimental set-up. Given that it is straightforward to set up a new experiment in the platform quickly, there was no pressing need

for this feature. However, this feature will be incorporated if future users of this platform request this ability.

3.6 Conclusion

In summary, this manuscript provides an open-source web-based platform for investigators in the motor learning field to easily create and conduct sequential finger tapping task studies. It allows researchers with no coding experience to design and manage their experiments completely online, with no setup requirements. The findings presented here establish the feasibility of obtaining valid results and pave the way to break the barriers to designing and conducting online motor learning experiments.

Chapter 4 Exploring Synergies in Brain-Machine Interfaces: Compression vs. Performance

4.1 Summary

Individuals with severe neurological injuries often rely on assistive technologies, but current methods have limitations in accurately decoding multi-degree-of-freedom (DoF) movements. Intracortical brain-machine interfaces (iBMIs) use neural signals to provide a more natural control method, but currently struggle with higher-DoF movements—something the brain handles effortlessly. It has been theorized that the brain simplifies high-DoF movement through muscle synergies, which link multiple muscles to function as a single unit. These synergies have been studied using dimensionality reduction techniques like principal component analysis (PCA), non-negative matrix factorization (NMF), and demixed PCA (dPCA) and successfully used to reduce noise and improve offline decoder stability in non-invasive applications. However, their effectiveness in improving decoding and generalizability for implanted recordings across varied tasks is unclear. Here, we evaluated if brain and muscle synergies can enhance iBMI performance in non-human primates performing a two-DoF finger task. Specifically, we tested if PCA, dPCA, and NMF could compress and denoise brain and muscle data and improve decoder generalization across tasks. Our results showed that while all methods effectively compressed data with minimal loss in decoding accuracy, none improved performance through denoising. Additionally, none of the methods enhanced generalization across tasks. These findings suggest that while dimensionality reduction can aid data compression, alone it may not reveal the "true" control space needed to improve decoder performance or generalizability. Further research is required to determine whether synergies are

the optimal control framework or if alternative approaches are required to enhance decoder robustness in iBMI applications.

4.2 Introduction

Individuals with serious neurological injuries (e.g., higher level spinal cord injury, severe stroke, amyotrophic lateral sclerosis [ALS]) often require assistive technologies to perform basic activities of daily living (Maresova et al., 2020). While simple devices such as sip-and-puff systems and eye trackers can assist in mobility and communication, these tools typically require unintuitive control strategies to use, do not allow the user to manage multiple degrees of freedom simultaneously, and limit their ability to interact naturally with their environment (Ware and Mikaelian, 1986; Collinger et al., 2013; Mougharbel et al., 2013). In response to these limitations, intracortical brain-machine interfaces (iBMIs) have become popular over the last couple of decades as a promising technology for individuals with paralysis (Lobel & Lee, 2014). iBMIs offer a more natural way of controlling external devices by directly decoding brain signals into behavioral commands, and have been successfully applied in various settings, including controlling robotic arms (Flesher et al., 2021; Hochberg et al., 2012) and decoding live speech (Card et al., 2024; Willett et al., 2021). Although impressive, iBMIs have enabled simultaneous control of only a few DoFs (Collinger et al., 2013; Nason et al., 2021; Deo et al., 2024; Shah et al., 2024; Willsey et al., 2024)—whereas the human brain can control hundreds of muscles at the same time for fluid motion.

One hypothesis for the brain's efficiency in managing multiple degrees of freedom is that the nervous system simplifies this complexity by functionally grouping muscles, called muscle synergies, to act as a single unit (d'Avella et al., 2003b; Holdefer & Miller, 2002; Overduin et al., 2015b; Tresch & Jarc, 2009). Although the anatomical basis of muscle synergies is debated

(Kutch & Valero-Cuevas, 2012b; Ranganathan et al., 2016; Ranganathan & Krishnan, 2012; Tresch & Jarc, 2009), the concept has proven practically useful for understanding the neural control of movement. To investigate these muscle synergies, researchers have used dimensionality reduction techniques (e.g., principal component analysis [PCA], non-negative matrix factorization [NMF]) on muscle data recorded across various tasks, including walking (Chvatal & Ting, 2012), reaching (d'Avella et al., 2006), and grasping (Overduin et al., 2008). Similarly, these findings have been extended to neural populations in the brain (i.e., brain synergies), suggesting that these dimensionality reduction techniques might reveal the underlying signals of neural circuits during different tasks (Degenhart et al., 2020; Gallego et al., 2017b). If muscle and brain synergies are indeed a fundamental component of the brain's control strategy, then they could serve as an optimal control space for brain-machine interfaces, leveraging the brain's inherent structure to facilitate effective control. However, this remains an open question: if synergies are more than an abstraction, they should ideally provide a perfect framework for iBMI control, allowing us to capitalize on the brain's natural control strategies for improving decoding performance.

Brain and muscle synergies obtained through dimensionality reduction techniques have already proven to be useful in many applications. They have reduced noise in noninvasive recordings (both artifact-induced and electrical noise (Islam et al., 2021; Winkler et al., 2011)), enhanced iBMI decoder stability across sessions (Degenhart et al., 2020; Karpowicz et al., 2022), and compressed the high-dimensional physiological data (Cozza et al., 2020). However, questions remain about their effectiveness in implanted recordings, their adaptability to new task contexts beyond their initial training, and whether the compression benefits of reducing the dimensionality of high-channel-count data vary between methods and physiological origin.

Addressing these issues and comparing different methods could enhance the functionality and clinical translatability of BMI and neuroprosthetic interfaces.

In this study, we examined the potential of brain and muscle synergies extracted from three popular dimensionality reduction techniques (PCA, NMF, and demixed PCA (Kobak et al., 2016) [dPCA]) to enhance intracortical BMI performance in predicting finger kinematics or forearm muscle activity in non-human primates (NHPs). We systematically assessed their ability to compress the data while maintaining performance, denoise the data by increasing the specificity of the recording, and facilitate generalization across different task contexts. We hypothesized that we could greatly reduce the dimensionality of brain and muscle data without losing performance and that removing some of the lower-variance dimensions would help remove noise and improve iBMI performance. Additionally, we hypothesized similar performance across methods in all experiments. Overall, we found that dimensionality reduction methods were useful for data compression, with dPCA showing the best performance, but did not show denoising or generalization benefits.

4.3 Materials and Methods

4.3.1 Experimental Design

4.3.1.1 Task and Data Acquisition

Two non-human primates (NHPs), monkeys N and W, were implanted with Utah microelectrode arrays in the pre-central gyrus at the arcuate sulcus, which is an anatomic landmark for ‘hand area’ of motor cortex in NHPs. Monkey N was implanted with two 64-channel arrays, while Monkey W received a 96-channel array, and both recordings were limited to 96 channels. Additionally, Monkey N was also implanted with eight intramuscular bipolar

electromyographic (EMG) electrodes in the forearm, targeting muscles involved in finger and wrist movements (Flexor Carpi Radialis (FCR), Flexor Digitorum Profundus (three electrodes, denoted by FDP-distal [FDPd], proximal [FDPp], and FDP), Flexor Carpi Ulnaris (FCU, excluded due to disconnection from target), Extensor Carpi Radialis Brevis (ECRB), Extensor Indicis Propius (EIP), and Extensor Digitorum Communis (EDC); Figure 4.1A).

Monkeys were trained to perform a 2-degree-of-freedom (2-DoF) dexterous finger task. This task required them to flex and extend their index and middle-ring-small (MRS) fingers independently while targets were shown in colors over a virtual hand on a screen in front of the monkeys (screen next to monkey in Figure 4.1A). Targets were presented in a center-out manner: every two trials, the targets appeared on the center, which corresponds to a resting position between flexion and extension. Successful trial completion required that the monkey maintain its finger positions within the targets for a preset hold time of 750 ms. At the end of each trial, monkeys were rewarded with fruit juice by using a juicer machine that turned on a pump for 100 milliseconds (juicer artifact in Figure 2). Finger flexion was measured using a custom-designed manipulandum equipped with daily-calibrated bend sensors, as described previously (Nason et al., 2021b).

Additionally, monkeys performed three variations of the task: one in which the wrist was flexed by 23 degrees (“wrist context”), another in which a spring to resist flexion was added (“spring context”), and a final one that included both changes (“spring+wrist context”). These context changes allowed us to explore changes in brain and muscle activity due to the variations in task while keeping the finger kinematics the same, as shown previously (Mender et al., 2023). Data acquisition included real-time brain and finger movement data collection via xPC Target (Mathworks), ensuring millisecond precision. These data were later synchronized with EMG

data. The 96 brain channels were sampled at 30kHz and threshold crossings were detected by setting a threshold of -4.5 times the signal root-mean-square during an initial calibration recording. For brain data analyses, only channels with at least one threshold crossing per second on average across the recording session were used. For each of these channels, we extracted the spiking band power (power in the 300-1000 Hz band, SBP, (Nason et al., 2020a)) feature, known for its low power consumption while maintaining decoding performance. The SBP and finger movements were then averaged into 20ms bins (Figure 1B, right). On the other hand, the EMG data processing pipeline involved downsampling the original signal to 2000Hz using an 800Hz anti-aliasing filter, followed by a band-pass filter (100 to 500Hz), signal rectification, and a final low-pass filter at 6Hz to extract the envelopes ("Envelope Extraction" in Figure 1B, left) (Gallego et al., 2018b; Mender et al., 2023). The processed EMG was then normalized to the peak activation of each day and the resulting signal is referred to as the muscle activity throughout the paper. The binned SBP and kinematics, as well as the muscle activity were used for all analyses, as opposed to the raw data for each.

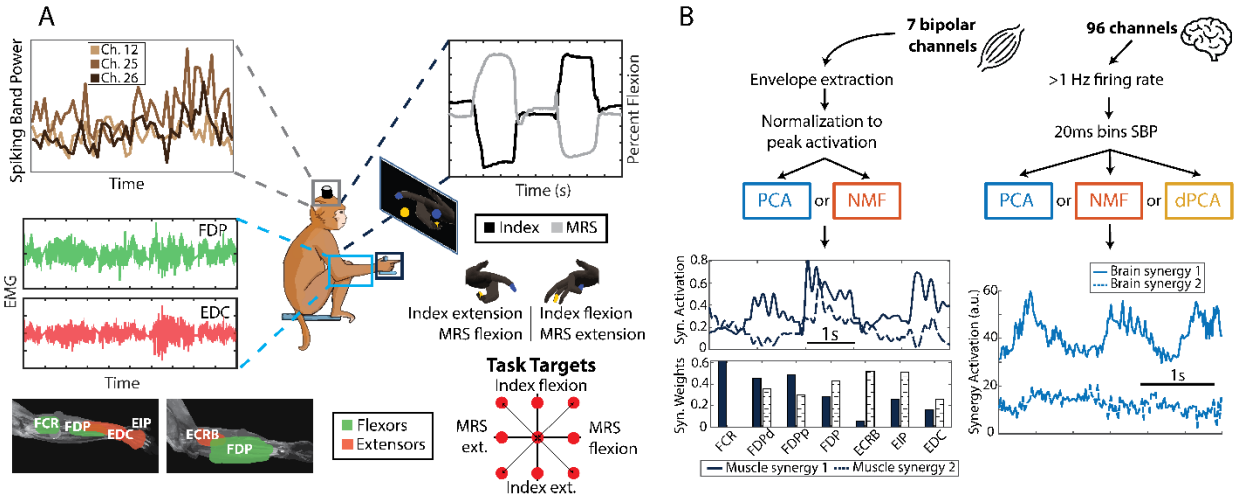


Figure 4.1: **Monkey task and data processing pipeline**. (A) Monkeys N and W were trained to do a 2-DoF dexterous finger task, in which they had to move the index and the middle-ring-small (MRS) fingers independently to targets shown on the screen (center diagram). While they did the task, we recorded finger position and velocity (top right plot), spiking band power of the brain channels (top left plot) and, for Monkey N, muscle activity (bottom left plot). (B) Diagram showing the data processing pipeline for the muscle (left) and brain (right) data. The 7

bipolar muscle channels are processed to extract envelopes and then normalized to the peak activation from each day. Then, we applied PCA and NMF, resulting in a set of muscle synergy activations (top plot) and synergy weights (bottom plot). The 96 brain channels are filtered to extract only those with at least 1Hz firing rate on average, and then transformed using the SBP in 20ms bins. PCA, NMF, and dPCA are then applied, which results in a set of brain synergy activations. **Abbreviations**: MRS = middle-ring-small fingers; EMG = electromyography; PCA = principal component analysis; NMF = non-negative matrix factorization; dPCA = demixed PCA; SBP = spiking band power; FDP = flexor digitorum profundus; EDC = extensor digitorum communis; FCR = flexor carpi radialis; ECRB = extensor carpi radialis brevis; EIP = extensor indicis propius.

4.3.1.2 Dimensionality Reduction Methods

We tested the application of different dimensionality reduction methods to the brain and muscle data. These methods work by taking an N-dimensional time series (such as the spiking band power of 96 channels throughout an experiment; N = 96 in that case) and extracting M (M <= N) synergies. These synergies can usually be described with two components: the synergy activations and the synergy weights (Figure 1B, bottom). The synergy activations are M-dimensional time series representing the corresponding synergy activity. In PCA, these activations are referred to as the scores. The synergy weights, on the other hand, represent the relative scaling of each of the original N dimensions for the corresponding synergy. In PCA, these are commonly referred to as the loadings.

We applied three dimensionality reduction methods to the data: PCA (Abdi & Williams, 2010), NMF (Rabbi et al., 2020), and dPCA (Kobak et al., 2016). All three were applied to brain activity, but only PCA and NMF were applied to the muscle activity, as dPCA's main benefit is extracting movement-relevant components from brain data (Kobak et al., 2016). All three methods identify low-dimensional components that can reconstruct the original data as best as possible, but they differ in the specific loss functions and their restrictions. First, PCA defines a decoder/encoder matrix D and tries to minimize the reconstruction error of the data X in the following form:

$$arg\ \min_{D} ||X - D^T DX||^2$$

In this study, we denominated DX as the synergy activations (a.k.a. scores) and D^T as the synergy weights (a.k.a. loadings). The D matrix can be solved analytically, making PCA decomposition a fast process. NMF, on the other hand, imposes a non-negative constraint on the weights (W) and the activation (H), and minimizes the reconstruction error of the data X in the following form:

$$arg \min_{W \geq 0, H \geq 0} ||X - WH||^2$$

NMF does not have an analytical solution; rather, numerical methods are needed to solve for W and H, making it a more demanding computation than for PCA. NMF is frequently used because it may better represent physiological signals due to its non-negative constraint (Rabbi et al., 2020). Finally, dPCA is similar to PCA, but it is supervised (i.e., has access to the behavior data for training) and also relaxes the restriction that the encoder and decoder transformations must be the same, which allows for selection of principal components in the brain activity most related to pre-specified task parameters. Like PCA, the dPCA matrices can also be solved analytically, using a similar objective function which accounts of task parameters, seen below:

$$arg \min_{F,D} \sum_{\phi} ||X_\phi - F_\phi D_\phi X||^2$$

Where ϕ represents the task parameter (i.e., target position or time), F_ϕ is the encoder matrix for that parameter, D_ϕ is the decoder matrix for that parameter, and X_ϕ is the trial-averaged data separated with regards to that parameter. In this study, we used two task parameters to separate the data: target position (nine possible values given by all combinations of flex, extend, and rest for each finger group; see Task Targets in Figure 4.1A) and time (i.e., a condition-independent parameter), both of which are common choices when using dPCA (Kobak et al., 2016). The synergy activation for dPCA was determined by computing $D_\phi X$, with

D_ϕ being the decoder matrix related to the target location, and X the non-parameter-separated data.

4.3.2 Statistical Analyses

4.3.2.1 Signal Quality Assessment

To assess the quality of the recorded physiological signals, we determined whether they were related to finger movements during the task. This tested the predictive ability of brain activity to estimate muscle activity and finger kinematics and of muscle activity to predict finger kinematics. For three days of 500 trials each with Monkey N, we trained three ridge-regression-with-history models (Collinger, Wodlinger, et al., 2013): one to predict muscle activity from brain activity, another to predict finger velocity from brain activity, and the last to predict finger velocity from muscle activity. For two days of 500 trials each with Monkey W, since only brain and kinematics data were available, we trained a single model predicting finger velocity from brain activity. All models had a total of 10 bins of history, equivalent to 200ms. Additionally, to study movement discrimination ability in the recorded signals, we trained linear classifiers (support vector machines with a linear kernel, *fitcecoc* in Matlab) that distinguished between nine different task targets (all combinations between flex, extend, and rest for each finger group; Figure 4.1A). The performance of the ridge regression models was measured with the Pearson correlation coefficient between the ground truth and the prediction, and the classifier's performance with the average accuracy across targets.

4.3.2.2 Relation Between Synergies and Movement

After assessing the quality of the signals, we applied dimensionality reduction methods to the brain and muscle data, and explored whether those tools could advance our understanding of the relationship between brain and muscle synergies with movement. Unlike muscle data, which

has a more direct and well-understood link to movement (e.g., FDPp is involved in the flexion of the index finger), brain signals can represent complex neural activity that doesn't always correspond in a straightforward way to intended actions, adding a layer of ambiguity and complexity in decoding. We quantified this brain and muscle interpretability hypothesis by computing the Pearson correlation coefficients between each muscle and brain channel with each kinematic output: index position and velocity and MRS position and velocity. Higher correlations between kinematics and muscle activity, than between kinematics and brain activity, may mean a more direct relationship between muscle activity and kinematics.

Then, we studied the relationship between brain and muscle and their respective synergies. We first determined the number of brain and muscle synergies needed to achieve at least 90% VAF for PCA (Frère & Hug, 2012; Hug et al., 2010; Steele et al., 2015; Turpin et al., 2021). Given a set of muscle and brain synergies for each method, we computed the Pearson correlation coefficient between every combination of brain channels and synergies with every combination of muscle channels and synergies. For example, for brain channel 1, we computed the correlation with all seven muscle channels, all PCA muscle synergy activations, and all NMF muscle synergy activations. If the correlations with the muscle channels are higher overall than with the muscle synergies, it may suggest that muscle activity is better represented in brain channel 1 than muscle synergies. To determine the significance threshold for correlations, we ran 1000 permutations of the signals for each pair of the computed correlations. We computed the thresholds with an alpha level of 5% and corrected for the number of comparisons using the Bonferroni correction. Finally, we organized the correlations into four groups: (1) brain channels vs. muscle channels, (2) brain channels vs. muscle synergies, (3) brain synergies vs. muscle

channels, and (4) brain synergies vs. muscle synergies. We compared the overall correlations between groups using the *cocor* package in R.

4.3.2.3 Compression Experiment

Next, we tested whether PCA, NMF, and dPCA were useful in compressing brain and muscle features when predicting muscle activity and kinematics. First, we computed the variance accounted for (VAF), as described in (Gallego et al., 2018b), for each method with both brain and muscle data, with varying numbers of synergies. Then, we trained ridge regression models that took varying numbers of brain or muscle synergies as input and predicted muscle activity and kinematics (when using brain synergies) or only kinematics (when using muscle synergies). We measured their performance by computing the Pearson correlation coefficient between the predictions and the ground truth. Additionally, we trained ridge regression models using the full high-dimensional brain or muscle data as input and compared their performance in predicting muscle activity and kinematics to that of ridge regressions trained with synergies. Given that the performance of these regressions may not necessarily exhibit a strictly monotonic increase, we determined for each the number of synergies necessary to achieve at least 95% of the correlation given by the full-data-trained model by fitting the following exponential function to the results:

$$corr(x) = k_1 + k_2 \cdot (1 + e^{-k_3 x})$$

Where *x* is the number of synergies used, *corr* represents the Pearson correlation coefficient, and $k_1$through $k_3$were fit using gradient-based optimization. After fitting the model, we found the *x* such that *corr(x)* went over 95% of the maximum value ($k_1 + k_2$). Then, we expressed the compression capabilities of each method for each input as the ratio between the total number of dimensions and this *x*. We tested the compression on three days for Monkey N and two days for Monkey W.

4.3.2.4 Denoising Experiment

Previous research has shown that removing dimensions of brain data (in modalities such as EEG or MEG) can help in decoding by eliminating noise (Islam et al., 2021; Winkler et al., 2011) (a.k.a., denoising). In this case, noise is defined as the electrical activity not related to the behavior of interest (e.g., movement artifacts, power line interference, or neural activity not related to behavior). We are interested in studying whether extracting synergies and then using them for decoding can help in denoising our intracortical and intramuscular data. We tested this hypothesis by training multiple models that evaluated muscle activity and kinematics prediction from brain and muscle synergies and compared them to the baselines of using all the available data. If the dimensionality reduction methods can denoise the data, then we would see better predictions coming from a reduced number of synergies than using all the data available. On the other hand, if the methods cannot denoise the data, then removing information, in the form of removing higher-order synergies, will reduce performance. To determine this, we performed the following analysis. We trained eight ridge regressions with different inputs and outputs (

Table 4.1, schematic in Figure 4.4A). Then, we compared the results of reducing the dimensionality using PCA (brain and muscles), NMF (brain and muscles), and dPCA (only brain), versus just using the full dimensional data. Note that for the models that predict muscle synergy activation rather than muscle activity, an additional layer was added that projects the output into muscle space using the appropriate muscle synergy weights, to compare the predictions across models in muscle space. To measure the performance of these models, we used the Pearson correlation coefficient between the predictions and the ground truth output. We tested denoising on three days for Monkey N, and two days for Monkey W.

Table 4.1: Ridge regressions for denoising analyses

Number	Input	Output
1	Brain activity	Muscle activity
2	Brain activity	Finger Velocity
3	Brain activity	Muscle syn. activation
4	Brain syn. activation	Muscle activity
5	Brain syn. activation	Finger Velocity
6	Brain syn. activation	Muscle syn. activation
7	Muscle activity	Finger Velocity
8	Muscle syn. activation	Finger Velocity

4.3.2.5 Generalization Experiment

Finally, previous studies have shown that latent spaces extracted with dimensionality reduction methods are stable across days (Gallego et al., 2018b), and we wanted to test whether they could also help when generalizing across task contexts. We had Monkey N perform the finger task in four different contexts (normal, spring, wrist, and spring+wrist) sequentially and then trained ridge regression models to predict finger velocity and muscle activity for each non-normal context. Off-context models refer to those trained using 100% of the data in the normal context and then tested using the last 20% of the data in each non-normal context (Figure 4.5A). On-context models were used as a baseline and refer to those trained using the first 80% of the data in each non-normal context and then tested on the last 20% of the data in the same context. Similarly to what we did for the denoising experiment, we trained off-context models to predict finger velocity and muscle activity from the full brain and muscle data (models 1, 2, and 7 in

Table 4.1), which are referred to as the baseline on-context models. Additionally, we trained off-context models to predict finger velocity and muscle activity from brain synergies and through muscle synergies (models 3, 4, 5, 6, and 8 in

Table 4.1; Figure 4.5A). We measured each model's performance using the Pearson correlation coefficient between the predictions and the ground truth and compared the performance between on-context and off-context models, as well as between models trained on synergies and models trained on the full data. We tested generalization on one day for Monkey N containing all four task contexts sequentially.

4.4 Results

4.4.1 Signal Quality Assessment

As an initial step, we assessed whether the recorded physiological signals were related to the monkey finger movements during the task. The recorded EMG in Monkey N visibly modulated activity with finger movements (Figure 4.2A). For example, FDP, a flexor muscle targeting middle-ring-small (MRS) flexion, increased activity when moving fingers towards flexion (Figure 4.2A, top), while EDC, an extensor muscle targeting MRS extension, increased activity when extending MRS (Figure 4.2A, bottom). Additionally, when using all EMG channels, the muscle activity significantly predicted finger velocity using a ridge regression with history (correlation coefficient CC = 0.60±0.08 across days; $p < 1E\text{-}4$; Figure 4.2B, right). Finally, we could use the muscle activity to accurately classify the correct kinematic target 97.64%±1.17% of the time (average across n=3 days; chance level 28.2%, nine total targets; Figure 4.2C).

Similarly, we wanted to verify the intracortical data reflected finger movements too. We recorded intracortical spiking band power (SBP; Nason et al., 2020) from 96 channels in the hand area of the motor cortex for two monkeys: N and W. The "active" channels for each, measured as those that had at least one spike per second on average, are shown in Table 4.2.

With Monkey N, the SBP predicted finger velocities with a slightly lower correlation (CC = 0.53±0.09; $p < 1E-4$; Figure 4.2B) compared to the predictions from muscle activity. Monkey N's SBP could also significantly predict muscle activity (CC=0.71±0.05; $p < 1E-19$) and accurately classify the task target for each trial (acc = 91.3%±4.6%; Figure 4.2C). Monkey W's SBP showed similar predictive capabilities compared to Monkey N, significantly predicting finger velocities (CC = 0.58±0.08; $p = 1E-3$; Figure 4.2D) and detecting the correct kinematic target (acc = 86.2%±0.9%; Figure 4.2C). Overall, these results show that the recorded physiological signals were related to the monkey finger movement during the task.

Table 4.2: Number of active intracortical channels per day and monkey

	Monkey N (total 96)	Monkey W (total 96)
Day 1	53	53
Day 2	37	29
Day 3	36	N/A

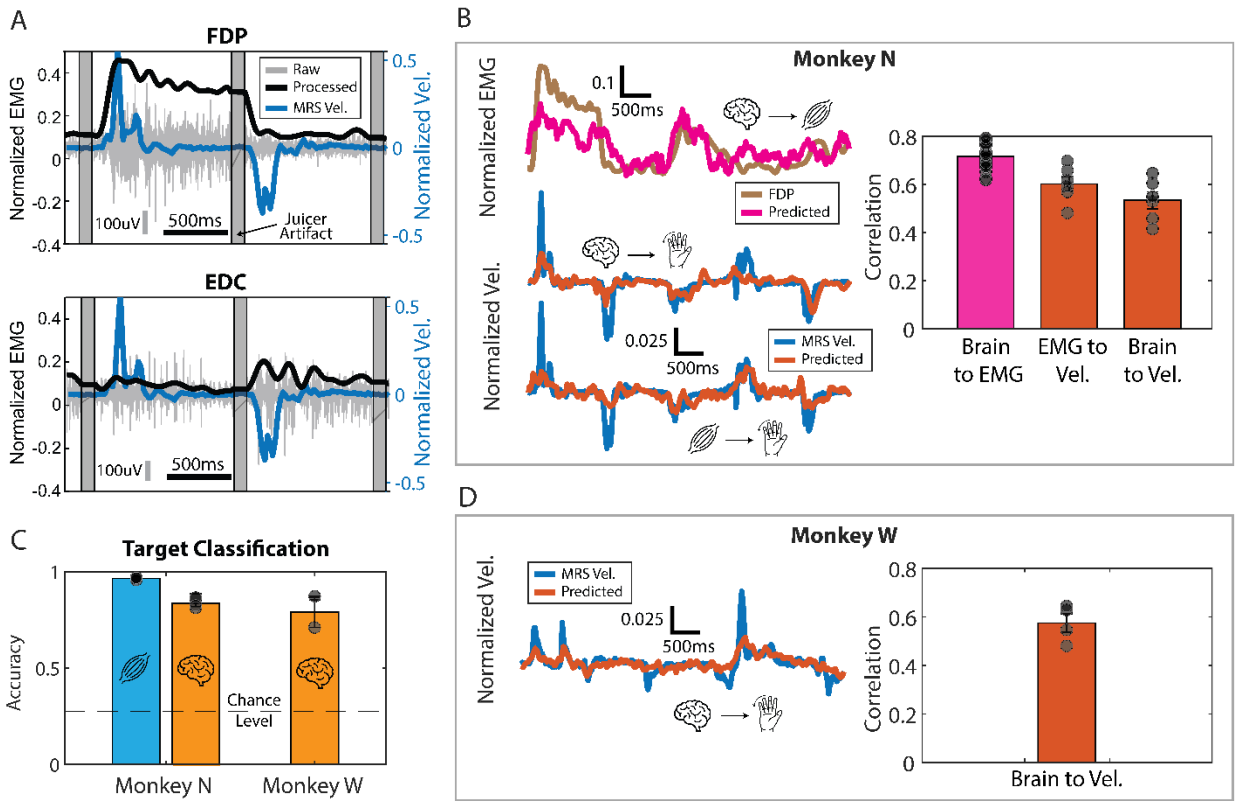


Figure 4.2: **Assessment of the quality of physiological signals**. (A) Monkey N's raw and processed EMG, together with MRS finger velocity for two example trials. Gray vertical bars represent the removed juicer artifact *(see Methods)* at the beginning of each trial. Top plot shows FDP, a muscle targeting MRS flexion, and the bottom shows EDC, a muscle targeting MRS extension. Vertical gray line shows scale of 100uV for the raw EMG and horizontal black line shows scale for 500ms. (B) Example traces of Monkey N's processed FDP EMG predicted from brain activity (top) and finger velocity predicted from brain activity (middle) and processed EMG (bottom). Bar plot on the right shows aggregated correlations corresponding to three days of Monkey N data. *Black* dots *in the left bar* represent the predictions for *each* muscle channel (*seven dots for each day*). *Bar plots in middle and right bar represent the velocity predictions for each finger group (two dots for each day).* (C) Accuracy of classifying targets using EMG (muscle symbol) and brain activity (brain symbol) for both monkeys across days. Segmented line shows chance level. Aggregated results correspond to three days of Monkey N and two days of Monkey W. (D) Same as in B but for Monkey W. Trace shows example of MRS velocity and the prediction when using the brain data. Bar plot shows aggregated results across 2 days of *M*onkey W data. **Abbreviations:** FDP = flexor digitorum profundus; EDC = extensor digitorum communis; MRS = middle-ring-small fingers; Vel. = velocity; EMG = electromyography.

4.4.2 Relation Between Synergies and Movement

Having shown that the recorded signals were, in fact, related to the monkey finger movements, we were next interested in exploring whether the dimensionality reduction methods could reveal distinct relationships between synergies and movements, particularly in the muscle data. Across all three days and methods (PCA, NMF), the extracted muscle synergies

consistently exhibited an explainable structure closely related to the task. Specifically, two synergies were predominantly composed of either flexor or extensor muscles (e.g., synergies 1 and 2 in Figure 4.3A, respectively), reflecting the flexion-extension nature of the center-out task performed by the monkeys. The third synergy was more broadly distributed among muscles, even those involved in opposing movement patterns, suggesting a supportive role in movement execution. This follows from a more direct relationship between muscles and movement (e.g., FDP is known to be involved in finger flexion) and was quantified by computing the correlations between individual muscle channels and specific movements (Figure 4.3B). In contrast, the individual brain channels are believed to have a more obscure relationship with movement, which hinders a similar interpretation to the muscle synergies in the context of the task, and was further validated by the lower correlations with movement, compared to the muscle synergies (Figure 4.3B).

This relationship between brain and movement may be obscure due to an intermediate modulation through brain and muscle synergies. We explored this hypothesis by investigating the relationship between brain and muscles directly, or through synergies. We computed the correlations between all possible pairs of brain and muscle channels (Nx7 pairs for each day, where N corresponds to brain channels in Table 4.2), as well as all pairs of synergies across days and methods (Figure 4.3C) and found generally low (CC <= 0.4) but significant correlations on most pairs: 93% for brain to EMG, 92% for brain to muscle synergies, 72% for brain synergies to EMG, and 71% for brain synergies to muscle synergies. Interestingly, we also found significantly higher correlations between brain and muscle activity than between all other pairings: higher than brain to muscle synergies, $p<1E-20$; higher than brain synergies to EMG, $p < 1E-20$; higher than brain synergies to muscle synergies, $p < 1E-20$. The stronger relationship

between brain and muscles than between using any type of synergy as an intermediate step raises the question as to whether synergies can actually be more useful for decoding than using all the available data, or at least whether they can simply help compress the data more effectively, as has been suggested previously (Cozza et al., 2020; Tang et al., 2021).

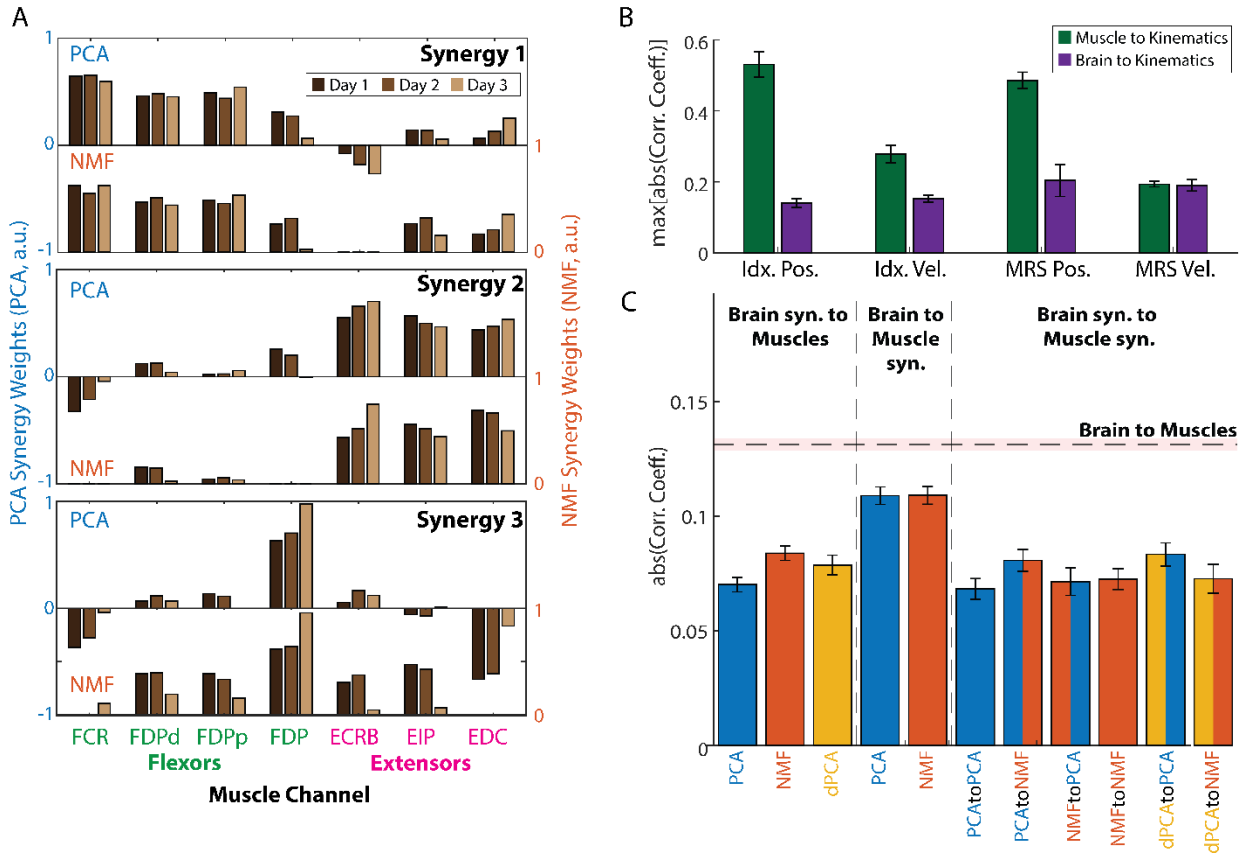


Figure 4.3: **Analysis of brain and muscle synergy relationships for monkey N**. (A) Muscle synergy weights extracted using PCA (top bar plot in each window) and NMF (bottom bar plot) across days. Synergy weights were consistent across days and methods and revealed a structure closely related to the task. (B) Barplot showing the average across days of the maximum absolute correlation coefficient between all muscle channels and all brain channels with the four kinematic outputs: index and MRS position, index and MRS velocity. The brain channels had a weaker direct relationship with the kinematic outputs, which hinders an interpretation of the brain synergies similar to that shown in A. (C) Average of the absolute correlation coefficient between all possible pairs of brain and muscle synergy activations, broken down by the method(s) used. The *black* segmented line shows the average of the absolute correlation coefficient between all *pairs of* brain and muscle channels, while the shading represents the standard error of the mean (SEM). All error bars represent the SEM as well. **Abbreviations**: PCA = principal component analysis; NMF = non-negative matrix factorization; FCR = flexor carpi radialis; FDP = flexor digitorum profundus; FDPd = FDP distal; FDPp = FDP proximal; ECRB = extensor carpi radialis brevis; EIP = extensor

indicis propius; EDC = extensor digitorum communis; Idx = index finger; MRS = middle-ring-small fingers; Pos. = position; Vel. = velocity; Syn. = synergy; Corr. Coef. = Pearson correlation coefficient; a.u. = arbitrary units.

4.4.3 Compression and Denoising Experiments

The ability to compress data effectively is useful by itself, as it can greatly lower power requirements and enable simpler decoding models (Casson et al., 2010; Thies & Alimohammad, 2019). Here, we first compared the different dimensionality reduction methods' abilities in compressing brain and muscle data in two ways: we first measured the VAF and studied how the total explained variance changed when increasing the number of synergies. Then, we evaluated compressibility end-to-end by measuring the performance of decoders trained on varying numbers of synergies as inputs (Figure 4.4A). In terms of VAF, three out of seven (~43%) muscle synergies consistently explained >90% of the variance, across all days and methods (Figure 4.4C, right). For brain synergies, ~54% of the total dimensions were needed on average to explain 90% of the variance (Figure 4.4C, left) for monkey N, while ~51% were needed for Monkey W (Figure 4.4E, left). Note that dPCA is extracted from the trial-averaged data and thus cannot explain the full variance of the data, even with all possible synergies. In terms of end-to-end compression ability, we found that all three methods were good at compressing brain data, achieving 95% of the maximum prediction correlation with a relatively small percentage of the total dimensions (Figure 4.4, B and E), resulting in an average compression ratio of 3.43:1 across monkeys and outputs (Table 4.3). When compressing muscle data, PCA and NMF had a smaller effect than on the brain data, achieving an average compression ratio across days of 1.75:1 (Table 4.3).

Table 4.3: Average compression ratios across methods and monkeys.

Prediction Type	Monkey	PCA	NMF	dPCA

Brain to Velocity	N	2.23:1	2.23:1	4.83:1
	W	2.90: 1	2.90:1	4.14:1
Brain To EMG	N	2.90:1	2.90:1	5.80:1
EMG to Velocity	N	1.75:1	1.75:1	N/A

Given the promising results in compressing brain and muscle data, we next examined whether dimensionality reduction could also enhance prediction accuracy by reducing noise in the data, i.e., whether it could "denoise" brain and muscle data. Specifically, we were interested in determining if fewer synergies, representing a compressed and potentially less noisy version of the original data, could lead to better predictions compared to using all dimensions. We trained models either using the full dataset or on a reduced set with varying number of synergies (see Methods) and compared their prediction performance. Our results showed that, on both monkeys, none of the dimensionality reduction models outperformed the baseline of using the full data (Figure 4.4B, D, and E; Supp. Figure C.1). When predicting muscle activity (Figure 4.4B, top row and heatmap in D; Supp. Figure C.1) and finger velocity (Figure 4.4B, bottom row and Figure 4.4E right), we observed a similar trend: as the number of synergies increased (and thus also the VAF), the prediction performance improved, but never above the baseline (indicated by the black lines on the line plots, 100% in heatmap). This lower performance was consistent across dimensionality reduction methods, with PCA and NMF showing nearly identical behavior and dPCA generally performing better with fewer synergies.

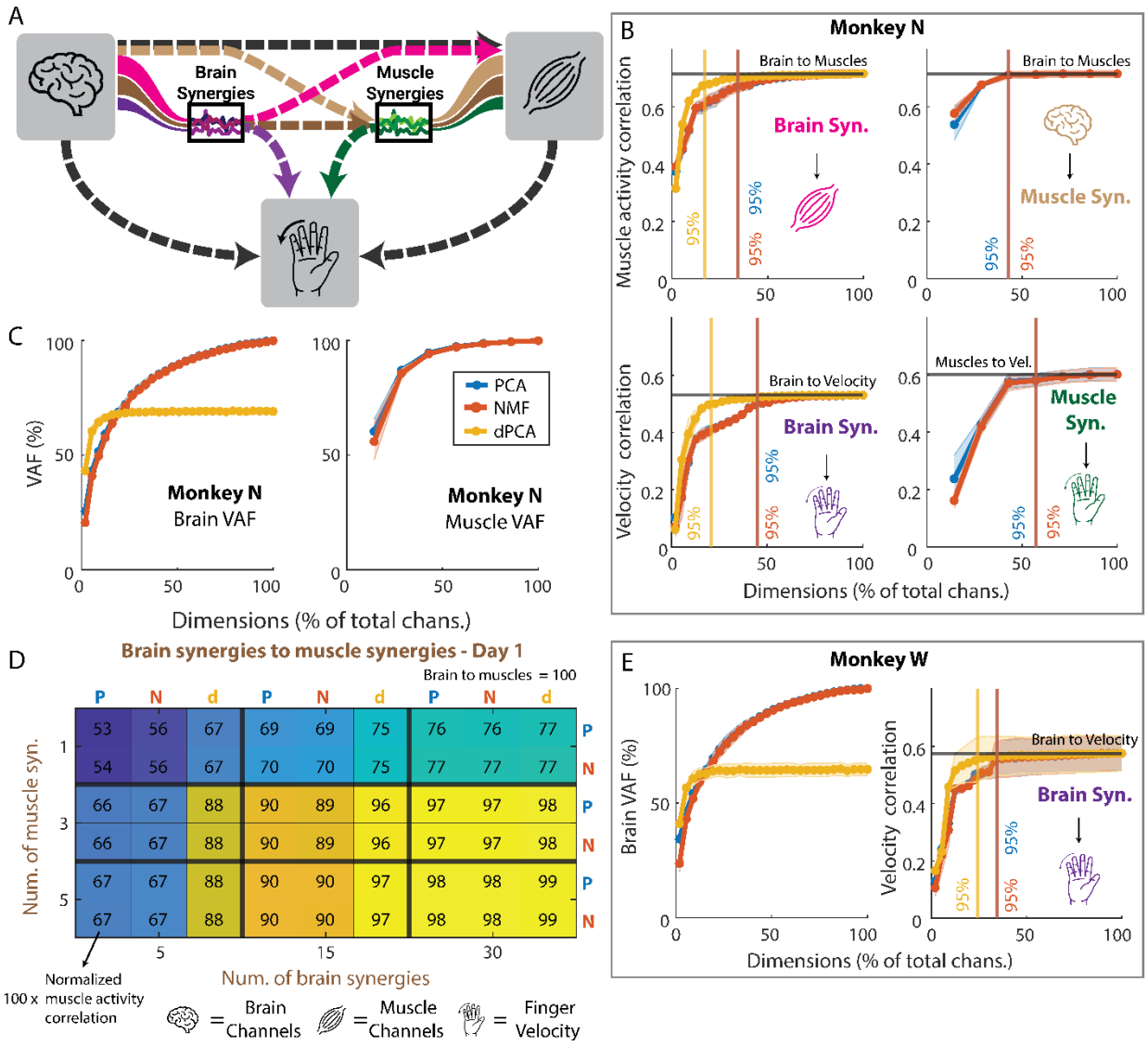


Figure 4.4: **Compression and denoising analyses**. (A) Diagram explaining the connections that were tested. Each line represents a different model. For example, the pink line represents a model that was trained with brain synergy activations as the input, and with muscle activity as the output. The three icons represent the brain channels, the muscle activity, and the finger velocity, respectively. The two black lines represent the baselines, shown in black in the rest of the plots. (B, top) Correlation coefficient of the prediction of muscle activity from brain synergies (left) and from the brain channels through muscle synergies (right). The x axis represents what percent of the total dimensions were used to make the predictions. The black lines represent the result of using all brain channels to predict all muscle activity. The colors of the icons are as in A. The vertical lines represent the percent of total dimensions that are needed for each method to achieve 95% of the maximum correlation. The shading for each line represents the standard error of the mean across days. (B, bottom) Correlation coefficient of the prediction of finger velocity from brain synergies (left) and from muscle synergies (right). (C) Average variance accounted for as a function of the percentage of total brain (left) and muscle (right) channels for monkey N. (D) *Sample h*eatmap showing, *for day 1*, the correlation of predicting muscle activity through brain and muscle synergies, normalized to the result of predicting muscle activity directly from brain channels. A value of 100 would mean that it matches the *baseline* perfectly. P represents PCA, N represents NMF, and d represents dPCA. (E) Same as B and C but for

Monkey W. **Abbreviations**: PCA = principal component analysis; NMF = non-negative matrix factorization; dPCA = demixed PCA; VAF = variance accounted for; Syn. = synergy; Vel. = velocity.

4.4.4 Generalization Experiment

Previous research suggests that even if synergies cannot help with denoising, they may be stable across tasks (Gallego et al., 2018b), which may enable higher generalization across tasks than using all data. We tested this hypothesis by having Monkey N perform the same task under different contexts: normal, with a spring resisting flexion, a 23-degree change in wrist angle, and both wrist and spring changes simultaneously. We trained decoders based on the full data or a set of synergies for both brain and muscles and predicted finger velocities, muscle activity, and muscle activity through muscle synergies. Off-context models were trained on normal trials, while on-context models were trained in the context they were tested on (see Methods, Figure 4.5A). Overall, we found that, regardless of context, as the number of synergies increased, the correlation with finger velocities and muscle activity increased but did not surpass correlations of the off-context models trained on full data (Figure 4.5B-F; Supp. Figure C.2). Therefore, there does not seem to be any obvious performance enhancing advantages of using synergies to tackle the problem of decoding generalization. The only exception is shown on Figure 4.5F, where NMF demonstrated a modest improvement in correlation in the spring+wrist context, compared to using the full dataset off-context (1-10%), but this effect was not observed in the other contexts (Supp. Figure C.2).

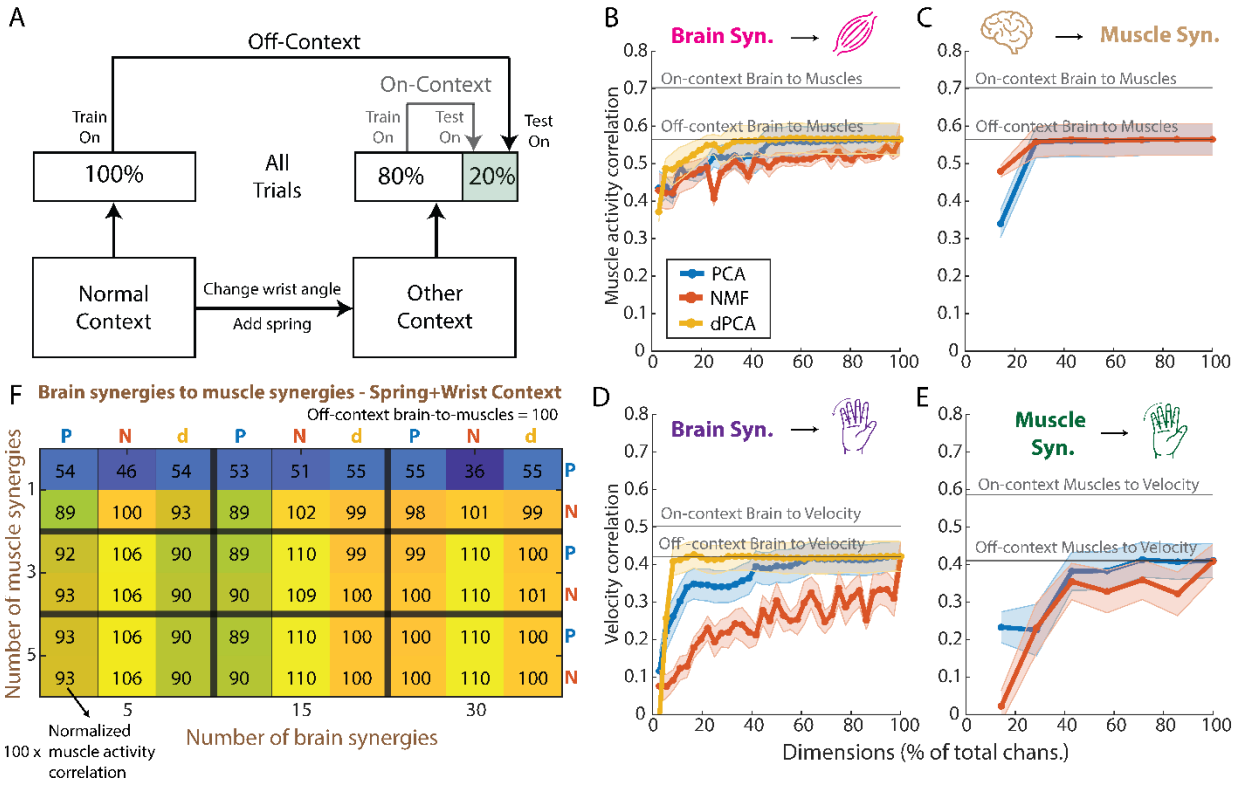


Figure 4.5: **Generalization analyses for Monkey N**. (A) Diagram showing how the on-context and off-context models differ: on-context models are trained on each non-normal context (spring, wrist, spring+wrist) and tested on the same context. Off-context models are trained on the normal context and then tested on each of the non-normal contexts. (B) Average across non-normal contexts of the Pearson correlation coefficient of the prediction of muscle activity from all numbers of brain synergies up to the total dimension of the brain data. The two black lines represent the baselines for on-context and off-context, when predicting the muscle activity from the full brain data. (C) Same as in B but for the prediction of muscle activity from the full brain data and through muscle synergies (see Methods). (D) Same as in B and C but now showing the prediction of finger velocity from brain synergies. (E) Same as in B, C, and D, but now for the prediction of finger velocity from muscle synergies. (F) Heatmap for the spring+wrist context, like the one shown in Figure 4.4D, showing the correlation coefficient of the prediction of muscle activity from brain synergies and through muscle synergies, normalized to the off-context baseline prediction. A value over 100 means a better prediction than the baseline off-context model that uses all brain data and predicts muscle activity directly. **Abbreviations**: PCA = principal component analysis; NMF = non-negative matrix factorization; dPCA = demixed PCA; Syn. = synergy.

4.5 Discussion

In this study, we tested the ability of three popular dimensionality reduction methods to compress and denoise brain and muscle activity from implanted electrodes, as well as their capacity for generalizing to unseen contexts. Our results demonstrate that while dimensionality reduction methods can effectively compress neural and muscle data, they do not improve decoding accuracy compared to using all available data or help generalize across tasks. Notably, while PCA and NMF performed comparably in compression, dPCA, a supervised approach,

generally showed better performance with fewer synergies, suggesting it may be a preferable approach when compressing for decoding models. Overall, these findings indicate that while synergies provide efficient data compression, they do not help data denoising for better performance than using the full dataset.

Brain and muscle synergies have been suggested in the field of neuroprosthetics as a practical way of solving different issues in performance. For example, (Degenhart et al., 2020) showed that brain synergies can be used to maintain BMI decoder performance across days by transforming each new day's brain activity to the synergy space. On the muscle side, (Ajiboye & Weir, 2009) showed that muscle synergies can be used to discriminate between postures of the American Sign Language and proposed their use as a control input for prosthetic devices. Here, we show that even though muscle synergies can be easily interpreted in terms of their kinematic meaning (in our case, two synergies for finger flexion, plus another for extension, Figure 4.3, A), the correlations between the brain and muscle activations were still always higher than between the brain and muscle synergy activations (Figure 4.3, C). These higher correlations suggest that a prosthetic control strategy that uses muscle synergies rather than muscle activation for control may be less related to brain activity, which can be undesirable when developing control strategies that are easy and natural to use.

Additionally, we showed that synergies extracted using dimensionality reduction techniques were not useful for denoising brain or muscle data from invasive electrodes. Synergies also did not appear to be useful in most cases for generalizing decoders for the same data across contexts. If these techniques could remove unrelated signals ("noise") from the data, then we would have seen better performance when projecting the data into a lower dimension than when using the full data. This denoising phenomenon has been shown in other studies

(Costa-García et al., 2023; Damon et al., 2013; Islam et al., 2021; Niegowski et al., 2015; Winkler et al., 2011), mostly in the context of less invasive recording methods. In our study, however, we found that removing dimensions to the data almost always results in worse performance (Figure 4.4), which could be driven by the higher inherent signal-to-noise-ratio (SNR) of invasive methods versus their noninvasive counterparts.

Overall, if the dimensionality reduction methods improved the stability of the decoders for brain and muscle across tasks, then we would have seen better generalization performance when using these methods than when using the full data. The stability of the lower dimensional projections of brain data across tasks has been shown in previous studies (Gallego et al., 2018b), and others have shown that decoder stability across time can be achieved by reducing the dimensionality and then adjusting future data points. Here, however, we show that dimensionality reduction methods on their own did not help a decoder generalize better than when using the full data in most cases. The results of denoising and generalization collectively suggest that higher-dimensional components, which are typically considered to lack task-related information (Barradas et al., 2020), do contain relevant information for decoding.

Although we were not able to achieve improvements by using synergies for denoising and generalization, these methods were still a great tool for compression. We found that all dimensionality reduction methods achieved good results, showing high compressibility for brain data (3.48:1 for Monkey N, 3.31:1 for Monkey W) and muscle data (1.75:1 for Monkey N; Figure 4.4). These results suggest that the brain data can be greatly compressed with no loss in decoding performance, which can be very valuable for more complex neural-network-based models. The size of these models is usually dictated by the dimension of the feature vector (Guduru, 2006), and bigger models increase the risk of overfitting to the training data (Srivastava

et al., 2014) and require more power to run in real-time (Casson et al., 2010; Gabert et al., 2020). Models that use less power are going to be the key to translating brain-machine interfaces into everyday patient use. Thus, using a linear dimensionality reduction technique as a first layer, potentially even implemented in hardware, that transforms the input features into a lower dimensional space without losing performance can be a key addition to existing complex models. A similar process has been used in other applications to improve model size (Mladenić, 2006).

The three dimensionality reduction methods we tested, PCA, NMF, and dPCA differ in the loss function they minimize and the restrictions they impose: PCA restricts the decoder to be the transpose of the encoder; NMF restricts the activations and the weights to be positive; dPCA allows for different encoder and decoder but computes the transformations subject to a task parameter of interest. In our results, we found dPCA to do the best in terms of compression, denoising, and generalization: it achieved higher compression rates than PCA and NMF (5.32:1 vs 2.57:1 for Monkey N; 4.14:1 vs 2.90:1 for Monkey W) and performed the closest to the full data in the denoising and generalization experiments (Figure 4.4 and Figure 4.5). These better results likely follow from the ability of dPCA to have different decoder and encoder transformations, the fact that it is trained on a supervised manner, and the possibility of using task parameters (target locations, in our case) to extract transformations relevant to the task. .

This study has some limitations that should be acknowledged. First, we did not test the full range of dimensionality reduction techniques available. We focused on PCA and NMF, which are commonly applied to brain and muscle data, respectively, and included dPCA for brain data to leverage task-specific information. However, other methods, such as more advanced manifold learning techniques or neural network-based approaches (Abbaspourazad et al., 2023; Pandarinath et al., 2018), may offer different or improved insights into data structure

and decoding performance, as well as allow for potential non-linear readouts from synergies to behavior. Second, our muscle dataset was limited to a single monkey that had synchronized recordings of muscle activity, brain signals, and kinematics. Although these experiments provide a valuable foundation, the generalizability of our findings is limited by the small sample size. We acknowledge the challenges and costs associated with such experiments, and we aim to expand our study to include additional subjects in future research to strengthen the validity of our conclusions.

In conclusion, the results of this study show that dimensionality reduction methods can effectively compress high-dimensional physiological data for brain-machine interfaces and neuroprosthetics, achieving substantial reductions in data size while retaining predictive power. However, these methods showed limited utility for denoising or improving generalization, as none outperformed the baseline (i.e., using all available data) for predicting EMG or finger velocity. Although synergies approximated full data performance in cross-context tasks, there was no significant advantage in enhancing generalizability. These findings highlight both the value and the limitations of dimensionality reduction for neural data applications, particularly in tasks requiring compact yet informative representations.

Chapter 5 Exploring the Trade-Off Between Deep-Learning and Explainable Models For Brain-Machine Interfaces[3]

5.1 Summary

People with brain or spinal cord-related paralysis often need to rely on others for basic tasks, limiting their independence. A potential solution is brain-machine interfaces (BMIs), which could allow them to voluntarily control external devices (e.g., robotic arm) by decoding brain activity to movement commands. In the past decade, deep-learning decoders have achieved state-of-the-art results in most BMI applications, ranging from speech production to finger control. However, the 'black-box' nature of deep-learning decoders could lead to unexpected behaviors, resulting in major safety concerns in real-world physical control scenarios. In these applications, explainable but lower-performing decoders, such as the Kalman filter (KF), remain the norm. In this study, we designed a BMI decoder based on KalmanNet, an extension of the KF that augments its operation with recurrent neural networks to compute the Kalman gain. This results in a varying “trust” that shifts between inputs and dynamics. We used this algorithm to predict finger movements from the brain activity of two monkeys. We compared KalmanNet results offline (pre-recorded data, n=13 days) and online (real-time predictions, n=5 days) with a simple KF and two recent deep-learning algorithms: tcFNN (non-ReFIT version) and LSTM. KalmanNet achieved comparable or better results than other deep learning models in offline and online modes, relying on the dynamical model for stopping while depending more on neural

[3] A version of this chapter “Exploring the trade-off between deep-learning and explainable models for brain-machine interfaces” was accepted for publication at NeurIPS 2024 (Cubillos, Revach, et al., 2024)

inputs for initiating movements. We further validated this mechanism by implementing a heteroscedastic KF that used the same strategy, and it also approached state-of-the-art performance while remaining in the explainable domain of standard KFs. However, we also see two downsides to KalmanNet. KalmanNet shares the limited generalization ability of existing deep-learning decoders, and its usage of the KF as an inductive bias limits its performance in the presence of unseen noise distributions. Despite this trade-off, our analysis successfully integrates traditional controls and modern deep-learning approaches to motivate high-performing yet still explainable BMI designs.

5.2 Introduction

Millions of people worldwide live with a neurological or spinal cord condition that limits their ability to interact with the world (Armour et al., 2016). These patients have to rely on others for basic tasks such as eating, moving, or even communicating with their loved ones, which severely limits their independence. Existing assistive devices, such as sip-and-puff controllers or eye trackers, can take advantage of small residual movements–if available–but are limited in the degrees of freedom (DoF) they can control. Brain-machine interfaces (BMIs) have started gaining traction in the past decades as a viable alternative that can control multiple DoFs while taking advantage of the existing hardware for computation: the brain.

Brain-machine interfaces work by reading electrical activity from the brain, decoding the activity with a decoding algorithm, and then using the algorithm's output to interact with an external device. Similar to existing applications such as language and image generation (Ding et al., 2021; Jiao & Zhao, 2019; Kang & Park, 2020; Otter et al., 2020; Zhou et al., 2020) and understanding (Deng et al., 2013; Guo et al., 2016; Ranjan et al., 2018; Samek et al., 2017), deep learning has become the norm in decoding algorithms for BMIs. New BMI applications such as

handwriting detection from a person unable to move (Willett et al., 2021) and real-time speech production from a person unable to speak (Metzger et al., 2023; Moses et al., 2021; Willett et al., 2023) have been enabled recently through the use of deep learning algorithms. Additionally, many deep learning models have demonstrated state-of-the-art results in BMI applications that previously only used linear decoders, such as real-time prediction of finger movement (Costello et al., 2023; Nason et al., 2021a; Willsey et al., 2022) and cursor control (Deo et al., 2024a; Wu et al., 2002).

These promising results using deep learning algorithms are enabled, in part, by an increase in the complexity of the models (Aggarwal & others, 2018). This can make interpreting their structure and results very difficult (Linardatos et al., 2020; Oh et al., 2019; Samek et al., 2017). It is also difficult to predict the algorithm's behavior with out-of-distribution inputs. This complexity has caused deep learning algorithms to be seen as “black boxes,” limiting their practical application in more sensitive domains such as vehicular control or lower limb prosthetics. Interaction with the physical world inevitably involves safety concerns. There, conventional control approaches that use explainable linear models are the norm, usually involving Kalman filters (KFs) because they are not as likely to exhibit unexpected behaviors. In BMIs, KFs and explainable linear models are still regularly employed in applications where the safety of the user is in question, such as when moving a robotic arm (Wodlinger et al., 2014) or controlling a paralyzed arm through electrical stimulation (Ajiboye et al., 2017). Thus, researchers are confronted with the need to sacrifice potentially better decoder performance in the name of safety and explainability. Here we ask: can we have the best of both worlds?

In this paper, we examine the trade-offs between linear, explainable models and high-performing deep-learning approaches. To interpolate between these two regimes (model-based

and data-driven), we modify and retrain the recently proposed KalmanNet (Revach et al., 2022), which combines the long-standing KF with the abstractness of deep models by augmenting its operation with a set of flexible recurrent neural networks (RNNs). We demonstrate the effectiveness of KalmanNet in real-time BMI experiments where monkeys with brain implants perform a dexterous finger task and show that we can match or outperform two previously published algorithms: tcFNN (non-ReFIT version) (Willsey et al., 2022) and LSTM (Costello et al., 2023). By leveraging the explainable portion of KalmanNet and its velocity dependence, we explain how this approach works through abrupt changes in the relative trust of the brain or dynamical models. Under the injection of additional noise, however, we find that while KalmanNet can reject low-magnitude noise, the LSTM actually outperforms all other models for large-magnitude out-of-distribution noise. Finally, we tested the ability of the models to generalize across task contexts and found better generalization for the simple KF than for all deep-learning-based models.

5.3 Related Work

Our work builds on previous modifications of the KF for BMI decoding, as well as advancements in deep learning and model-based techniques. Modifications to the standard KF, such as adapting it to non-linear dynamics or observation models, are common for BMI decoding. For example, Li et al. (Z. Li et al., 2009) used an unscented KF with a quadratic neural tuning model, achieving superior performance over the traditional KF, while others have shown that online adaptation of the KF parameters can have significant impacts on performance (Dangi et al., 2011; Gilja et al., 2012; Tsui et al., 2009). Additionally, modifying the trust variable of the KF--the Kalman gain--has been used to improve the computational efficiency by pre-fixing the trust value (Malik et al., 2011). Our work further extends this concept by allowing the Kalman

gain to vary over time, enabling a more flexible trust with a deep RNN. Pure recurrent and feed-forward neural networks have shown promise in BMIs. Feed-forward networks with a temporal convolution layer have outperformed the KF in an online finger control task when using the ReFIT recalibration method (Willsey et al., 2022), while RNN approaches, such as (Sussillo et al., 2012) and (Costello et al., 2023) have achieved even better results in predicting hand and fingers position and velocities in primates. KalmanNet incorporates the benefits of these deep learning approaches with the knowledge of real-world dynamics to improve the explainability of the model while achieving similar performance. Thus, our work also aligns with broader efforts in model-based deep learning (Shlezinger et al., 2023), where domain knowledge is combined with deep learning to enhance performance and explainability. For example, physics-informed neural networks (Raissi et al., 2019) incorporate physical knowledge into the loss function as a regularizer. KalmanNet can be framed similarly, with real-world dynamics serving as a regularizer for the underlying deep neural networks.

5.4 Methods

5.4.1 Neural Decoders

Throughout this study, we compared the performance of four neural decoders, three of which have been studied before in a BMI context: the Kalman filter (KF) (Kalman, 1960), the temporal convolutional feed-forward network (tcFNN) (Willsey et al., 2022), and a decoder based on the long short-term memory recurrent neural network (LSTM) (Costello et al., 2023). Here, we additionally adapt a recent model, KalmanNet (KNet), (Revach et al., 2022), to the BMI setting. Each decoder took in processed neural features and predicted movement kinematics (positions and velocities of the fingers; see Appendix D: for more details).

5.4.1.1 Kalman Filter

The KF algorithm tracks a state from noisy observations using a state-space (SS) description of the dynamics and the relationship between the observations and the tracked state. These are modeled as linear functions assuming Gaussian noise (Durbin & Koopman, 2012). In our study, the KF tracks a state composed of the position and velocity of both finger groups (total length of four). Like all algorithms in this study, it uses the 50ms binned spiking band power (SBP) (Nason et al., 2020b) of the brain activity as observations of the state (Figure 5.1, B). The KF computes the level of trust in the dynamics versus the observations with a variable called the Kalman gain (Figure 5.1, C), and then uses it to merge both sources of information. When the linear Gaussian SS representation holds, the KF is known to achieve the optimal mean-squared error (MSE) (Durbin & Koopman, 2012). Equation (5.1) shows how the Kalman gain (**KG**) is used to output a kinematics prediction by interpolating between the prediction given by the dynamics and the one given by the brain activity: a high value of the Kalman gain up-weights the brain activity as the main source of information, while a low value ensures the dynamics are the dominant term. The trainable parameters for the KF correspond to the linear observation model (C), the linear dynamics model (A), and the noise covariances of the state and the observations.

$$\text{Prediction} = \text{Dynamics Prediction} \times (\text{I} - \boldsymbol{KG} \times \text{C}) + \text{Brain Activity} \times \boldsymbol{KG} \quad (5.1)$$

5.4.1.2 tcFNN

The temporal convolutional feed-forward neural network (tcFNN) (Temmar et al., 2024; Willsey et al., 2022), is a "pure" (data-driven) deep learning approach that uses a temporal convolutional layer on the input, four fully-connected linear layers, and a two-dimensional output to predict finger velocities. It has been shown to predict finger velocities better than linear approaches in offline and online BMI experiments (Temmar et al., 2024; Willsey et al., 2022). The specific tcFNN architecture used throughout this paper was presented in (Temmar et al.,

2024) (see Appendix D: for more details). Here, however, the ReFIT process (Gilja et al., 2012), which showed the best results for online experiments in (Willsey et al., 2022) was not included, to compare only the baseline performance across algorithms.

5.4.1.3 LSTM

The long short-term memory (LSTM) (Costello et al., 2023) network is a type of RNN that uses memory cells and gates to control what information to remember and what to forget. The LSTM has achieved state-of-the-art results in finger movement predictions, outperforming the tcFNN, the KF, and other linear and non-linear approaches (Costello et al., 2023). The specific LSTM architecture used throughout this paper was presented in (Costello et al., 2023).

5.4.1.4 KalmanNet

KalmanNet (Revach et al., 2022) is a tracking algorithm that uses a (possibly mismatched) SS representation combined with deep learning (Shlezinger et al., 2023); It uses the KF as an inductive bias, by incorporating a deep learning architecture based on RNNs that augments the Kalman gain computation (Figure 5.1, C). This inclusion of the RNN allows for a more flexible modulation of trust between dynamics and observations while maintaining the linear model used to model dynamics (evolution model, A) and the relationship between states and observations (observation model, C). It differs from the KF in that it does not need an explicit model of the noise covariance of dynamics or observations, nor does it need to track the state covariance through time. The parameters of KalmanNet are the SS model parameters (necessary for the KF) and the trainable parameters of the included RNNs. Here, we modified KalmanNet to use the same dynamics and observation models that have shown success in the KF in the past (Nason et al., 2021a; Vaskov et al., 2018), modified the loss function to account for

the difference in scales between positions and velocities, and optimized the hyperparameters of the network for predicting finger positions and velocities.

5.4.2 Task and Data Acquisition

Two non-human primates, Monkeys N and W, were implanted with microelectrode arrays in the hand area (precentral gyrus) of the right motor cortex of the brain. Monkey N was implanted with two 64-channel arrays and Monkey W with a single 96-channel array; in both cases, recordings were limited to the best 96 channels. We recorded the SBP (power in the 300-1000Hz band) on each channel, a useful and low-power feature for BMI applications (Nason et al., 2020b). The SBP was averaged in 50ms nonoverlapping bins for each channel. For training and testing, we only used the "active" channels (i.e., those with at least one threshold crossing per second on average, approximately 20-30 total channels each day) for KalmanNet. For all other decoders, we used all channels that had threshold crossings morphologically consistent with action potentials (Costello et al., 2023; Willsey et al., 2022) (approximately 60-70 total channels, which included the "active" channels).

Both monkeys were trained to do a 2-degree-of-freedom (2-DoF) dexterous finger task in which they had to move the index and middle-ring-small (MRS) fingers independently to acquire targets shown on a screen (Figure 5.1, A). To successfully acquire targets, monkeys had to move their fingers to the required position and hold for a minimum hold time (750 ms). A new set of targets appeared on each trial, and the monkey had five seconds to acquire them; if the monkey could not acquire the targets during that time, the trial was labeled as unsuccessful. On *hand control* trials, finger kinematics were measured using resistive bend sensors and shown to the monkey in real-time using a custom visualization tool (MSMS (Davoodi & Loeb, 2011)). On *brain control* trials, the 50-ms binned SBP was fed in real-time to a decoder that predicted finger

kinematics and controlled the output visualization. The two modes–*hand control* and *brain control*–were almost identical, with the exception being the reduced hold time (500 ms) during brain control trials to lower their difficulty. Analyses of performance during hand and brain control trials are referred to as offline and online, respectively.

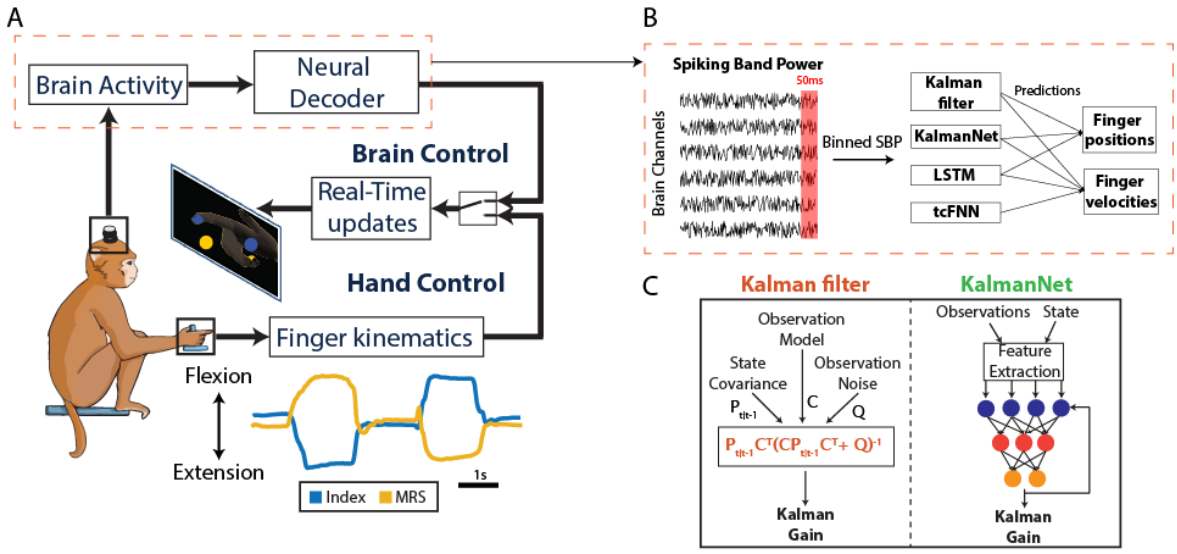


Figure 5.1: **Task and Neural Decoders.** (A) We trained a monkey to do a 2-DoF finger task (shown on screen) while brain activity and finger kinematics (index and middle-ring-small (MRS) traces shown below) are recorded. The monkey can do the task in hand control, using his hand, or in brain control. (B) In brain control, the SBP of each brain channel is extracted and binned every 50 milliseconds. Each neural decoder takes in a bin and predicts position and velocity, or only velocity. (C) The KF and KalmanNet differ in how they compute the Kalman gain: the KF uses the equation shown, while KalmanNet uses a set of RNNs.

5.4.3 Decoder Training and Testing

For offline decoder testing, decoders were trained in a supervised manner using a pre-recorded set of 400-500 hand-control trials from each day with synchronized brain activity and finger kinematics. The decoders were then tested on a different set of 100-500 hand-control trials from the same day to evaluate offline performance. We trained decoders daily, as micro-movements in the microelectrode array can cause large changes in the recorded neural activity across days (Perge et al., 2013). We trained and tested 13 different instances of each decoder overall, 8 using Monkey N's data, and 5 using Monkey W's data, and they were evaluated by

their ability to predict finger positions and velocities (for KF, KalmanNet, and LSTM) or just finger velocities (for the tcFNN).

When running online (real-time trials), decoders were trained on a similar initial set of 400-500 hand-control trials and then tested in pairs each day in an ABA manner: decoder 1 for a set of trials, decoder 2 for another set, and then back to decoder 1 for another set, to account for possible shifts in signals or monkey behavior. All decoders were tested using a similar number of trials in each set. KalmanNet was compared online with the KF on four days, with the tcFNN on two days (during which the KF was also run), and with the LSTM on one day. All days on which the monkey attempted to do online trials with KalmanNet and at least one more decoder were included in our analysis. Only Monkey N was used for online experiments.

5.4.4 Performance Metrics

We used different metrics to determine decoder performance when tested offline versus online. When testing a decoder offline, we have access to the ground-truth finger kinematics and thus can directly measure prediction accuracy. We used the Pearson correlation coefficient and the MSE between the ground truth and the predictions and then averaged across fingers to get a single value for each position and velocity. To compare offline performance across models, we used a combination of frequentist and Bayesian statistical methods. First, paired-sample t-tests on the MSEs as well as in the Fisher Z-transformed correlation coefficients (Meng & Rubin, 1992) determined whether the difference was significant ($\alpha = 0.05$) under the null hypothesis assumption. Second, we computed the Bayes factor (B_{01}) to determine the ratios of likelihoods between the null and alternative hypotheses.

When testing a decoder online, without access to ground-truth finger kinematics, we can simply measure how well the monkey completed the trials. We used four metrics for online

performance: success rate, trial times, path efficiency, and throughput. The success rate is the ratio of the number of trials successfully completed to the number of trials attempted. Trial time is the time taken to acquire both finger targets minus the hold time: higher trial times mean a slower or more erratic movement. The path efficiency measures the smoothness of the movements by computing the ratio between the distance between the start position and the target and the distance traveled by the monkey when completing the trial: higher path efficiencies mean smoother paths. Finally, the information throughput (measured in bits/s) accounts for each trial's difficulty (distance between start and target) and completion time (Willsey et al., 2022), with a higher throughput representing faster target acquisitions. Note that all metrics except the success rate were computed only with successful trials. Since there is some variability in performance across days due to micro-movements of the electrode array (Perge et al., 2013), we normalized all the metrics--except success rate--from each online day by the average performance of KalmanNet during that day, to be able to aggregate each decoder's performance across testing days.

5.4.5 KalmanNet as a Non-linear Trust System

To understand the mechanism used by KalmanNet to accurately predict finger kinematics, we inspected the network's output, which corresponds to the Kalman gain matrix in the KF. For easier interpretation, we computed the Frobenius norm of KalmanNet's Kalman gain and compared it to the norm of the predicted velocities. As a metric of the relationship, we used the Pearson correlation coefficient between the Frobenius norm of the Kalman gain and the vector norm of the velocities over time across all 13 offline days tested for both monkeys (n=8 for Monkey N, n=5 for Monkey W).

Additionally, to replicate KalmanNet's observed correlation between Kalman gain and velocity in the regular KF, we modified the KF to have heteroscedastic process noise. We refer to this model as the heteroscedastic Kalman filter (HKF). The process noise determines how noisy we believe the dynamics to be: with noisier dynamics, the KF will trust observations more (higher Kalman gain), and with less noisy dynamics, the KF will trust the dynamics more (lower Kalman gain). Thus, to force the Kalman gain to covary with the velocity, we made the process noise, specifically the parts related to the velocities, increase linearly with the norm of the ground truth velocity (see Appendix D: for more details). Note that this model needs the ground truth velocity to properly vary the process noise, which makes it unsuitable for online experiments.

5.4.6 Robustness to Real-World Operation

To test the models' offline robustness to real-world operation, we first simulated noise spikes on the data. This may happen in a real-world scenario with, for example, movement artifacts or loose electrical connections, and we investigated whether this would have unsafe consequences when controlling a physical device, such as big spikes in the velocity predictions. The noise spikes were simulated by adding a fixed magnitude to all neural channels of the testing dataset for a given number of 50-millisecond bins. We varied the magnitude from 0.1 to 100 times the standard deviation of the training set SBP and varied the duration from one to five 50-millisecond bins. While this may seem like a large amount of noise, for BMI applications, it is possible to have unexpected mV level noise added to μV level neural signals. The noise was injected randomly for 5% of the total time steps on each testing day, and we tested it offline across 13 different days for both monkeys. We measured the impact of the noise in the models

with the MSE of the velocity normalized by the baseline velocity MSE (no noise added to the testing dataset).

Additionally, we tested the decoders' ability to generalize to unseen task contexts, a key property for decoders to work in real-life scenarios rather than just in highly constrained research settings (Mender et al., 2023; Temmar et al., 2024). On six different days, we modified the task by having the wrist flexed by 30 degrees and adding a spring to resist flexion and facilitate extension (Mender et al., 2023). Then, we trained off-context and on-context decoders. Off-context decoders were trained on a normal set of trials (no wrist or spring modifications) and then tested on the last 20% of the spring+wrist context trials. On-context decoders were trained on the first 80% of the spring+wrist context trials and then tested on the last 20% (see Figure 5.6, A). We evaluated the generalization performance of each algorithm by measuring the percentage increase in velocity MSE from the on-context to the off-context decoders.

5.5 Results

5.5.1 Offline and Online Performance

To determine how well KalmanNet could fit the data compared to state-of-the-art deep learning techniques, we first assessed its offline performance in predicting finger kinematics from brain data. We evaluated KalmanNet over n=13 days across the two monkeys, comparing it with the KF and the deep learning based tcFNN and LSTM models.

The LSTM was the highest-performing approach in a previous study (Costello et al., 2023). In terms of correlation with velocity, which controls the visualization when running online, KalmanNet did not significantly differ from the LSTM ($p = 0.64$, $B_{01} = 3.3$) and had a significantly higher correlation than the KF ($p < 1E\text{-}7$, $B_{01} = 0.048$) and the tcFNN ($p < 0.001$, $B_{01} = 0.062$; Figure 5.2, B). All approaches had similar correlations with position, other than the

tcFNN, which is a velocity-only approach. Regarding velocity MSE, there was also no significant difference between KalmanNet and the LSTM (p = 0.72, B_{01} = 3.4) or between KalmanNet and the tcFNN (p = 0.29, B_{01} = 2.1). However, KalmanNet significantly outperformed the KF (p < 0.01, B_{01} = 0.14). In terms of position, there were no significant differences in correlation (KNet vs. LSTM, p = 0.32, B_{01} = 2.4; KNet vs KF, p = 0.93, B_{01} = 3.5; LSTM vs. KF, p = 0.08, B_{01} = 1.1) or MSE (KNet vs. LSTM, p = 0.26, B_{01} = 2.0; KNet vs KF, p = 0.77, B_{01} = 3.5; LSTM vs. KF, p = 0.1, B_{01} = 1.1), although KalmanNet had higher variance in MSE. Overall, it is important to note that the absence of p-values below the significance level for position and velocity correlations and MSE between the LSTM and KalmanNet should not be interpreted as definitive evidence that the two models are equivalent. Instead, it indicates that the data did not provide strong enough evidence to conclude that the models were different. However, the Bayes factor (B_{01}) values between 2 and 3.4 suggest that the models may have comparable performance, with the null hypothesis (models are equivalent) being at least twice as likely as the alternative hypothesis (models are different). This suggests a higher likelihood that any observed differences are not substantial, though it remains essential to consider the variability and sample size when interpreting these results.

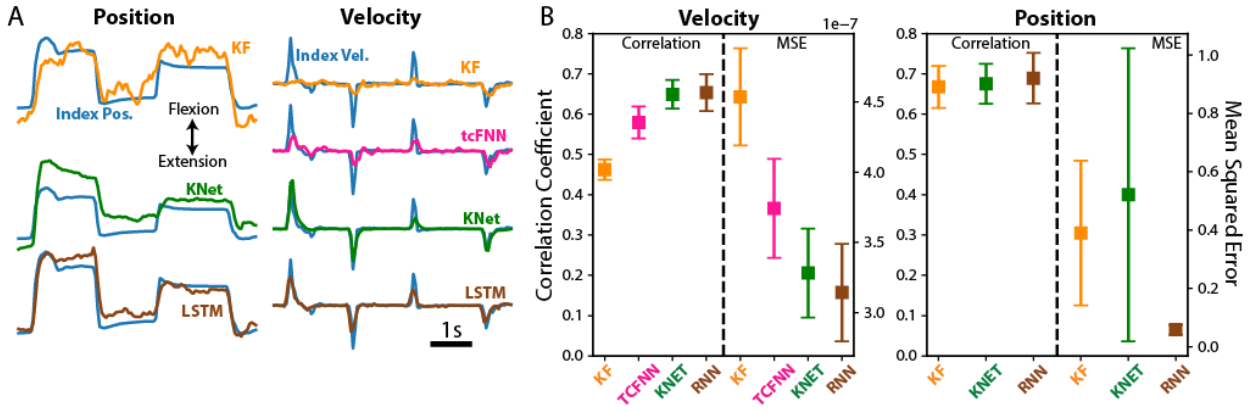


Figure 5.2: **Offline Performance.** (A) Traces of ground truth index position and velocity (in blue) versus the predictions from each neural decoder. Note that tcFNN only predicts finger velocity. (B) Velocity (left) and position

(right) performance in terms of correlation and MSE for each neural decoder. Square markers and error bars denote the mean and the standard error of the mean, respectively. Tested across $n=13$ days from both monkeys

However, because this is a motor control application, it is important to evaluate performance online, with a user in the loop. Therefore, we tested all algorithms online across five sessions (number of trials: KF = 601, tcFNN = 576, KNet = 2801, LSTM = 393) and compared their performance in terms of success rate, throughput, trial times, and path efficiencies. Figure 5.3 (right) shows the performance of all approaches normalized to the performance of KalmanNet (except for the success rate). We found that KalmanNet outperformed the KF and tcFNN across all metrics (Figure 5.3), allowing the monkey to perform faster and more accurate trials. Compared to the LSTM, KalmanNet achieved lower throughputs ($p < 1E-16$) and higher trial times ($p < 1E-5$), meaning slower trials overall. However, it had higher overall success rates (97% vs. 90%) and more efficient and smoother paths ($p < 1E-7$; Figure 5.3), which is consistent with the inclusion of KalmanNet's dynamics model. The online example traces tell a similar story, with the KF and tcFNN showing more oscillations in the path and more unsuccessful trials (Figure 5.3, A).

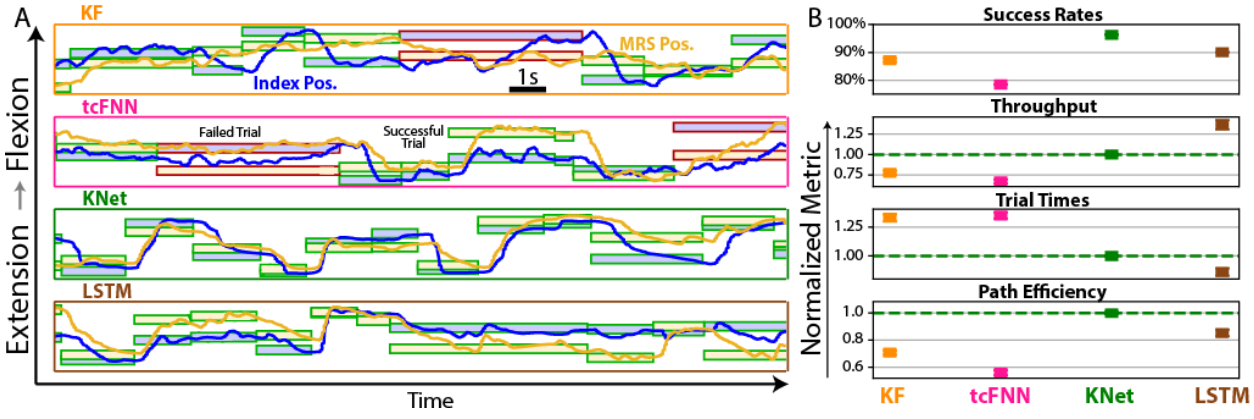


Figure 5.3: **Online Performance.** (A) Traces of the positions of the index (blue) and middle-ring-small (yellow) fingers during online control, across neural decoders. Blue-filled boxes represent index targets in the flexion-extension range, while yellow-filled boxes represent MRS targets. Green-outlined boxes represent successful trials, while red-outlined boxes show failed trials. (B) Online metrics of performance. Throughput, trial times, and path efficiencies are normalized to the corresponding KalmanNet value for each day. Tested across $n=5$ days from Monkey N.

5.5.2 Modulating Trust between Dynamics and Observations

Notably, KalmanNet can match or nearly match the highly nonlinear black-box LSTM performance. In KalmanNet, the structure of the linear KF is intact, such that the performance improvement comes from the RNNs used to compute the Kalman gain. This is the KF's metric of trust that determines how to interpolate between dynamics and observations (see Methods). To analyze this, we first inspected the regular KF's Kalman gain and observed that it converges to a fixed value in the first few seconds of a session (inset in Figure 5.4, A, orange line), in line with its known steady-state behavior (Durbin & Koopman, 2012). We then evaluated the KalmanNet gain computation during inference. Interestingly, we found that it correlated strongly with KalmanNet's output velocity, both offline (mean + standard error of the mean = 0.79 ± 0.02) and online (0.60 ± 0.02; Figure 5.4, center). This correlation suggests that KalmanNet behaved as a non-linear trust system, switching rapidly towards trusting the neural activity when the velocity was high and quickly switching to trusting the dynamics when it needed to stop.

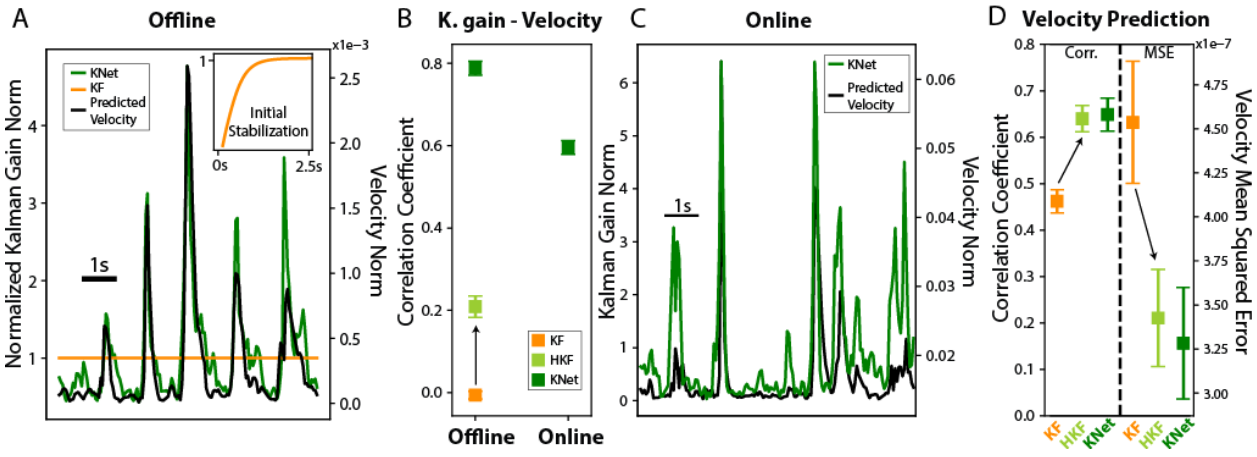


Figure 5.4: **KalmanNet as a Non-linear Trust System**. (A) Offline trace showing the norm of the Kalman gain for KalmanNet (in green) and for the KF (in orange), together with the predicted velocity for KalmanNet (black). The inset shows the first five seconds of trials, during which KF's Kalman gain converges to a fixed value~\cite[Ch. 4]{durbin2012time}. (B) Average correlations between Kalman gain and velocity for KalmanNet, HKF, and KF, for offline ($n=13$ days, across both monkeys) and online ($n=5$ days, only Monkey N) trials. (C) Online trace of the norm of the Kalman gain for KalmanNet (green), together with the predicted velocity (black). (D) Offline velocity prediction performance comparison between KF, HKF, and KalmanNet, in terms of correlation coefficient (left) and MSE (right). Arrows represent the addition of a heteroscedastic process noise to transform the KF into HKF. Tested across $n=13$ days and both monkeys.

To further explore whether this observed modulation of trust is the key to KalmanNet's improved performance, we used our knowledge of the true velocity to manipulate the noise covariance of the regular KF to achieve the same effect. Specifically, we created a new KF model with heteroscedastic process noise that increased with high velocities and decreased with low velocities, forcing the Kalman gain to covary with the output velocity (HKF model; Kalman gain correlation with velocity output across days and monkeys = 0.21 ± 0.03; Figure 5.4, B). We found that this modification made the new model perform as well as KalmanNet when predicting velocity, both in terms of correlation (HKF = 0.64 vs. KNet = 0.65, p = 0.57) and MSE (HKF = 3.4E-7 vs. KNet = 3.3E-7, p = 0.73), which showcases the benefits of the trust modulation mechanism observed in KalmanNet.

5.5.3 Robustness to Noise and Change in Task Context

The primary motivation for using a principled KF rather than a black box neural network is its safe operation, even under new or changing circumstances. The KF conventionally leverages physics-based SS models to improve kinematic predictions from noisy sensors. Therefore, we evaluated these approaches in the face of artificially-injected sensor noise of varying magnitudes and duration across channels (see Methods). This allowed us to study the effect on the output velocity under various real-world scenarios. Given the small magnitude of brain signals, noise artifacts can be much larger than the signal features of interest. We modeled those with extreme noise additions ~100 times the standard deviation in the training data. Surprisingly, KalmanNet had the largest increase in MSE associated with injected noise (by a factor of ~4300 for noise on three consecutive time bins, when the noise magnitude was 100x the std. dev.). It exhibited high-velocity spikes at the noise injection points (Figure 5.5A and Supp. Figure D.3). Perhaps surprisingly, the best performer was the LSTM, which changed very little

from the no-noise case. It showed just a ~2x increase in velocity MSE when injected with a noise level 100 times the training data's standard deviation. It is important to note that the noise injected was not present in the training data, so even the regular KF performed poorly. However, unmodeled noise is a common scenario in BMIs stemming from, for example, problems with the physical connection to the recording electrodes or electrode movement with respect to neurons (Perge et al., 2013).

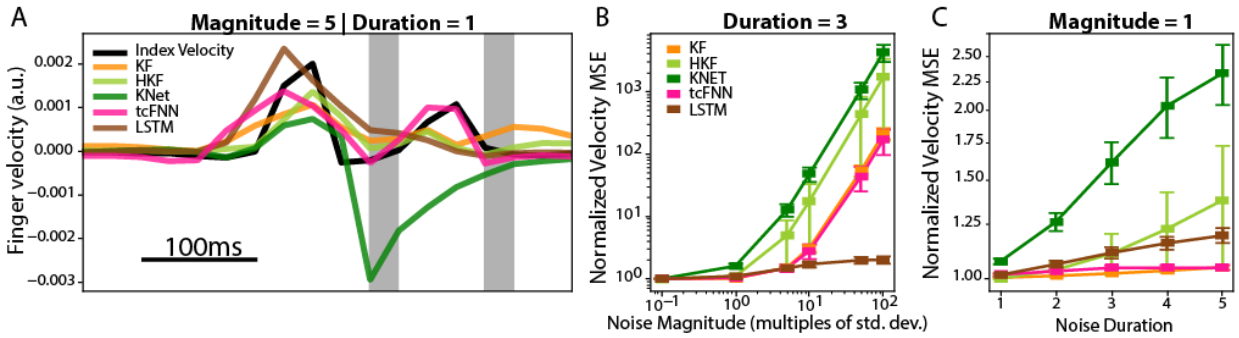


Figure 5.5: **Robustness to Injected Noise**. (A) Offline traces across neural decoders as compared to the ground truth index velocity (black). Grey columns represent noise injections of 5 times the standard deviation of the training dataset distribution, for a duration of one time bin. (B) Change in normalized velocity MSE as the noise magnitude changes, for a fixed noise duration of 3 time bins. (C) Change in normalized velocity MSE as the noise duration changes, for a fixed noise magnitude of 1x the training dataset distribution's standard deviation. Tested on 13 days for Monkeys N and W.}

Another potential benefit of a KF with a simple linear model is the ability to generalize to new contexts. Therefore, we evaluated all decoders by slightly varying the monkey task. We trained all models on the normal task context and then tested them in a new context (Mender et al., 2023; Temmar et al., 2024) (off-context performance), which contained a spring to resist flexion, as well as a 30-degree angle change in wrist position (Figure 5.6, A). We then compared their performance to training and testing in the new context (on-context performance). The normal KF showed the lowest drop in absolute performance (7.98% MSE), exhibiting good generalization to the new model but still showing the lowest absolute performance. On the other hand, all neural network models, including KalmanNet, showed substantially higher MSE increases due to the context shift (34.20% for tcFNN, 44.45% for KNet, 50.34% for LSTM).

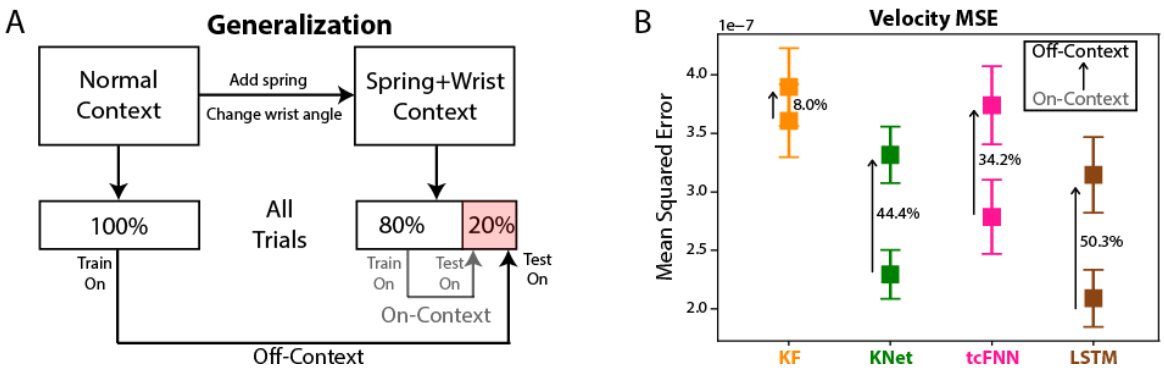


Figure 5.6: **Generalization Across Task Contexts**. (A) Illustration of how off-context and on-context decoders are trained. Monkeys performed a variation of the task that included a change in wrist angle and a spring that resisted flexion. Off-context decoders were trained on a normal set of trials from the same day and then tested on the last 20% of the trials of the spring+wrist context. On-context decoders were trained and tested on the same spring+wrist context. (B) Change in velocity MSE between on and off-context decoders for each decoder. Tested across n=6 days of Monkey N.

5.6 Discussion

In this paper, we explored the trade-offs between explainable linear models and deep-learning approaches to BMI decoding. We modified and retrained KalmanNet (Revach et al., 2022), which fuses deep learning techniques with the classic KF. We showed that KalmanNet can match or outperform existing state-of-the-art decoders in offline and online modalities. We also showed that the explainability of KalmanNet can be leveraged to improve a linear model up to the state-of-the-art standard but that it comes at the cost of greater error when faced with out-of-distribution noise.

KalmanNet approached the state-of-the-art performance offline, achieving comparable velocity and position predictions to the LSTM (Costello et al., 2023). This performance does not necessarily transfer to online trials, as has been studied previously (Cunningham et al., 2011; Deo et al., 2024a), and as we observed here with the tcFNN model. Offline, the tcFNN had significantly better velocity predictions than the KF, but online, it had overall worse success rates, throughput, trial times, and path efficiencies. This lower online performance may follow from the fact that the monkey is in the loop, reacting in real-time to the output of the decoder,

which may change the dynamics of the problem versus testing decoders offline (Cunningham et al., 2011; Deo et al., 2024a). Note, however, that we did not implement ReFIT (Gilja et al., 2012; Willsey et al., 2022) for tcFNN, a recalibration technique that can improve tcFNN online performance (Willsey et al., 2022); a similar method could potentially be developed for all decoders (albeit with potential differences in performance improvements), and thus we decided to compare all models with their baseline non-ReFIT performances (see Appendix D:). KalmanNet, on the other hand, was able to work well online and did similarly well to the LSTM: it acquired targets more slowly but had higher success rates as well as smoother paths, which may be more desirable for applications controlling physical devices. Moreover, while we currently use existing SS modeling that is learned separately from the filtering task, the fact that KalmanNet converts the KF into a machine-learning model indicates that one can potentially learn the observation model (C) and its computation of the Kalman gain. As demonstrated in (Buchnik et al., 2024), this could lead to both improved MSE performance and provide information on the usefulness of existing approaches for forming SS representations in BMI.

KalmanNet functioned as a non-linear trust system, quickly re-weighting the contributions to its predictions from either the model inputs or internal dynamics depending on the desired output. For stopping, since the dynamical model provides an explicit term for exponentially decreasing the output velocity (see Appendix D:), KalmanNet learned to trust the dynamical model. For fast movements, only the neural activity contains this information. Therefore, it chose to put more confidence in the neural inputs. We constructed a new linear model that enforced this observed behavior (HKF), and it matched offline state-of-the-art results, albeit assuming that the velocity level is known. This showcases some of the potential benefits of using a more explainable and, therefore, trustworthy model compared to a black-box decoder.

Additionally, the smart switching between dynamics and inputs is an interesting behavior that could be enforced even in black-box deep-learning decoders to potentially improve decoding. Similar approaches were recently explored for speech neuroprostheses, using a combination of a language model (which can be thought of as language "dynamics") and an RNN predicting phonemes from neural activity (Willett et al., 2023). One could modify that approach by, for example, having a model that learns to trust the neural activity for starting words or when faced with uncommon transitions between phonemes, and to trust the language model on more common phoneme transitions or for being silent. This approach would be similar to the behavior we observed in KalmanNet.

A disadvantage of combining a neural network with a Kalman filter is that the overall method carries over some limitations from both approaches. We found that when faced with out-of-distribution noise, KalmanNet had the highest increases in MSE, up to 1000x, when faced with noise 100x the original distribution's noise. This effect likely follows from the similar behavior exhibited by the KF, which increased its MSE 100x with the same noise magnitude. Nonetheless, the fact that KalmanNet preserves the operation of the KF indicates that it can be extended to outlier-robust variations of the KF, e.g., (Truzman et al., 2023), or alternatively, combined with hypernetworks for handling multiple noise levels (Ni et al., 2024). Interestingly, the pure RNN approach (the LSTM) had the lowest increase in error at high noise magnitudes. This result may be unexpected for control-focused researchers, who often choose linear approaches due to the perceived higher safety of the more explainable models.

In conclusion, the results shown here with KalmanNet suggest that we may not need to sacrifice explainability to achieve results in line with the state-of-the-art deep-learning models

and that using a mechanism based on gating trust between dynamics and observations may provide valuable insight for developing future models.

5.7 Acknowledgements

We thank Eric Kennedy for animal and experimental support and Professor Hans-Andrea Loeliger for helpful discussions. This work was supported by NSF GRFP 1841052, NSF EFRI BRAID 2223822, CASI award (1021865.01) from the Burroughs Wellcome Fund, NSF NCS 1926576, NIH R01-NS-105132-01-A1, NIH U41 NS129436 and the Agencia Nacional de Investigacion y Desarrollo (ANID) of Chile.

Chapter 6 Discussion

This Discussion explores the broader implications and future directions of the findings presented in this thesis. Section 6.1 focuses on how the results contribute to our understanding of motor control and motor learning, including insights into ankle impedance and the brain's reliance on dynamics for control. Section 6.2 examines the relevance of these findings for assistive technologies, highlighting their potential to improve device design, usability, and power efficiency while addressing challenges like explainability in deep learning models. Finally, Section 6.3 proposes future research directions, emphasizing the need for studies in less controlled environments and the development of decoding algorithms that integrate real-world knowledge to address limitations in current approaches.

6.1 Implications for Motor Control and Motor Learning

The findings of this thesis have important implications for our understanding of motor control. First, in Chapter 2 we found that measuring ankle impedance during standing and walking using perturbations was highly reliable, particularly when measuring stiffness and viscosity. These findings further validate the efforts of previous researchers, who have mapped the ankle’s impedance parameters throughout the gait cycle (Lee et al., 2016). This mapping helps us understand how the nervous system modulates mechanical impedance during standing and walking, Paired with other studies that have studied this process at the muscle level (Joshi et al., 2022), it can give us key insights into how this modulation assists in balance and stability.

The results from Chapter 5 may also provide additional insights into the mechanism behind motor control. We found that KalmanNet, an algorithm that combined deep learning with a principled approach, could achieve high performance in velocity decoding by relying on brain activity for moving fast and on physics for stopping. These findings support the idea that the primary motor cortex (M1) may prioritize computational simplicity in its generation of movement commands, which can be described as a form of 'lazy' control (Famié et al., 2022; Gaveau et al., 2016). When generating the commands to complete a certain trajectory, M1 may encode mostly high-velocity changes, and rely on the modeled dynamics of the body for stopping and maintaining steady-state positions (Famié et al., 2022; Gaveau et al., 2016; Mistry et al., 2005). When training KalmanNet, the algorithm had no explicit instruction or loss to rely more on brain activity or dynamics at any given time; it just naturally learned that the brain provided more reliable information during movement and that the modeled dynamics made for better information for stopping.

This idea of the brain encoding change better than it does steady states is not new. Researchers studying the primary somatosensory cortex (S1) have found similar effects when recording neural activity during touch. When touching a surface with a finger, the neural activity follows a stereotypical activity during which neurons fire fast at the beginning of the touch, then lower their activity during steady-state touch, and then fire fast again when the finger stops touching the surface (Okorokova et al., 2018; Westling & Johansson, 1987). This insight into how the brain encodes sensory information has been used to produce “biomimetic” patterns of stimulation that follow this fast-slow-fast curve to emulate a more natural sensation of touch through intracortical micro stimulation (ICMS) (George et al., 2019; Okorokova et al., 2018). Studies have shown that by using these biomimetic patterns of stimulation, the elicited sensations

are more natural and stronger, for a given amount of delivered electrical charge (Greenspon et al., 2024).

6.2 Implications for Assistive Technologies

The findings of this thesis have significant implications for the development of assistive technologies, particularly in terms of improving the design, functionality, and usability of prosthetic and rehabilitation devices. In Chapter 2, I show that we can reliably measure the ankle's stiffness and viscosity during standing and walking. These metrics can serve a double purpose. First, they can be used as a clinical metric for tracking progress in patients with stroke or other conditions that affect their motor control. Previous studies have shown differences in ankle stiffness and viscosity between healthy and stroke populations (Shorter et al., 2021), which suggests that those metrics can be used to determine the effectiveness of specific rehabilitation therapies. Second, these metrics provide insight into how the brain achieves accurate and efficient motor control during standing and walking. This has guided the development of new lower-limb prosthetic and exoskeleton control strategies that try to match the natural modulation of stiffness and viscosity during walking, and have shown to be a more natural and efficient control strategy (Azocar et al., 2020; Best et al., 2023).

These biomechanical insights that lead to potentially more natural control strategies are, however, hard to obtain for all human joints and all activities, which makes designing easy-to-use control paradigms challenging for all assistive devices. Fully understanding how humans learn new skills can help us design control paradigms that are easy to learn, but running experiments that give insight into this process can be challenging with existing tools, as many of these require technical expertise that most researchers in the motor learning field do not have. This is why, in Chapter 3, I introduced a new web application for motor learning researchers that

requires no technical expertise and instead allows them to design simple keypress experiments using a form-like interface. Using this new application, we gained new insights into how interlimb transfer learning works for key-pressing and replicated previous results showing that even short periods of rest can help learning. These two insights can be guiding principles for designing control strategies for assistive devices as well as training procedures for the patients using them.

Another issue that current assistive devices struggle with is their power requirements. The battery on microprocessor-based prosthetics and exoskeletons is not enough for many users to get them through whole days (e.g., 8-hour battery life for Ottobock's Empower foot). For implanted devices, there is an additional limitation, where too much power usage may heat the surrounding tissue above the safety limits (Wolf & Reichert, 2008). In neuroprosthetic devices, sampling from the sensors and using a model to decode useful control commands from them use most of the available power. Thus, if we can improve the power usage of the sampling and decoding step, we may make big steps toward usable and safe assistive devices. In Chapter 4, I show how dimensionality reduction methods such as principal component analysis (PCA), non-negative matrix factorization (NMF), and demixed PCA (dPCA) may be valuable compression tools for decoding. I showed that for brain and muscle data we can achieve compression ratios of up to 5.80:1, reducing the size of the sensor data by almost 6x. Thus, these tools can be a great addition to the signal processing pipeline, potentially reducing power consumption and thus increasing the usability and safety of these assistive devices.

Most of these devices, and part of the reason why they need so much power, have transitioned lately to using deep learning models for the prediction of intended movements given sensor information. Deep learning models have shown the best results on most applications

across domains (Card et al., 2024; Costello et al., 2023; Willett et al., 2021; Willsey et al., 2022), but suffer from an inherent problem of explainability: most of these models behave as "black boxes" where interpreting how the model is processing the input is unclear. The potential issue this brings is that unexpected inputs in real-world use may have unforeseen consequences, which can be severe if the patient strongly relies on the device. My work in Chapter 5 directly addresses this issue by using a model that combines deep learning with a more principled approach and achieves similar results to pure "black-box" models. These results also highlight two specific benefits of explainability. First, an explainable model that achieves good predictions given some sensor data can help us understand the inputs when its source is more obscure. For example, in my work in Chapter 5, we show that by analyzing the behavior of the model during predictions we could observe a reliance on the neural activity to produce high velocities and on the dynamics for stopping. This finding may help us understand better how the brain modulates activity to produce movement. Second, an explainable model that achieves good predictions can help us develop new models that are either simpler, and thus, subject to fewer assumptions, or overall better. In Chapter 5, I show how, by understanding that the key to KalmanNet's high performance was its modulation of trust dependent on desired velocity, we could build a fully linear model that achieved similar performance to a pure black box model. These findings suggest that we may not need to sacrifice explainability to achieve good results, and that explainability may have desirable side effects.

6.3 Future Directions

The research presented in this thesis opens the way for some exciting new avenues of exploration in human motor control and assistive devices. Two paths that are particularly interesting to me are moving from the lab to the wild world and decoding algorithms that use

knowledge about the world. Most of the experiments shown in this thesis were in highly controlled environments: in Chapter 2, users were instructed to land in a specific position while walking, in Chapters 4 and 5 the monkeys did the finger task while having most other degrees of freedom constrained. Controlled environments are key for testing specific hypotheses, and without them, it would have been impossible to advance as far as we have in terms of scientific achievements. However, when studying motor control, the act of constraining the environment may change how motor control is expressed and thus constrain the validity of the results. We observed this effect on the generalization results in Chapter 4: even the simple act of changing the angle of the wrist in which the monkeys performed the finger task made the decoder trained on normal trials significantly lose performance. In brain-computer interfaces this is a common issue: most experiments are run in highly controlled experiments, and thus their validity in the wild world may be limited. I believe that a necessary future direction is to study motor control in less constrained conditions, which will require ingenious experiment designs to properly test scientific hypotheses, especially in the realm of brain-machine interfaces.

The second future direction I propose is to design decoding algorithms that incorporate some knowledge about the world, to overcome some limitations with pure deep learning. When designing and training deep learning algorithms for any application, the key is to collect enough representative ground-truth data and then make the model complex enough to capture the complex relationships of the input. Many times, however, collecting enough data, as well as making sure it is representative of the task, is a challenging problem. In brain-machine interfaces, for example, the ground truth is collected by having a subject either do or imagine (depending on the level of injury) certain movements, while the movement and the neural activity are recorded. Then, the model is trained to match the ground-truth movement given the

recorded neural activity. This approach comes with some inherent problems. If the subject is only imagining the movement, it is difficult to ascertain that the movement prompted is exactly what they are imagining. Additionally, the collected data may not be representative of the neural activity when the decoding algorithm starts working in a closed loop: previous work has shown that the neural dynamics change when running in a closed loop (Cunningham et al., 2011; Deo et al., 2024a).

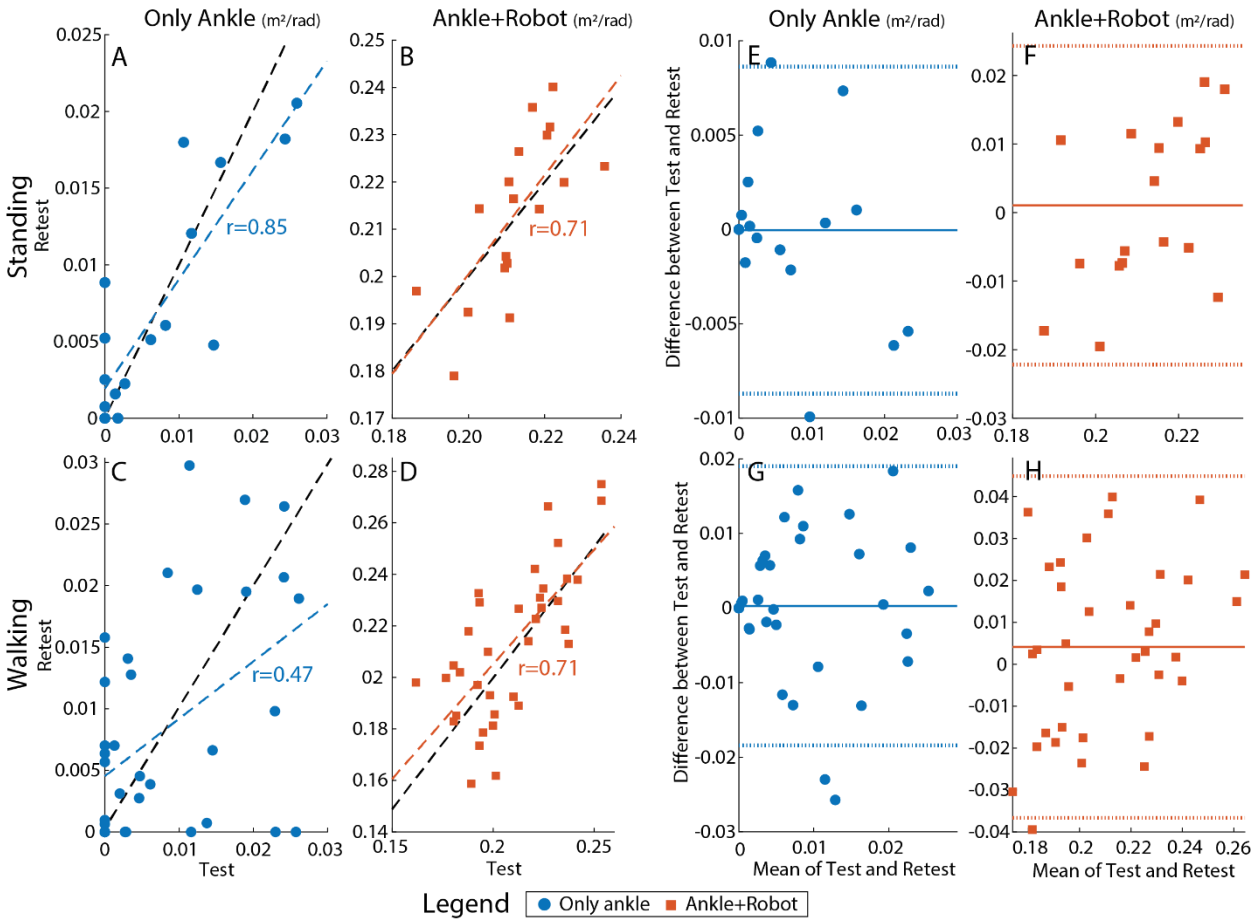
In Chapter 5, I showed that KalmanNet could match the performance of pure deep-learning algorithms. It did this by incorporating knowledge about the world, in the form of a model of the state dynamics through time. The deep learning part of KalmanNet only decided whether to look at the state dynamics or the brain. This approach, I believe, can be valuable for all applications in which collecting representative ground truth data is challenging. The knowledge about the world can act as a regularizer that overcomes some of the limitations the training data may have. Additionally, incorporating knowledge about the world in a model that is meant to interact with the world seems appropriate, especially since other disciplines such as physics have studied it extensively and can lend the appropriate models. This is the basis for physics-informed neural networks (Raissi et al., 2019), and I believe that we can go even further, as demonstrated with KalmanNet in Chapter 5.

Overall, I believe that the next 10-20 years will show incredible advancements in assistive technology. The ability to record from an increasing number of neurons through electrophysiology and imaging, together with an exponential increase in computing capabilities will cause a boom in studies of the nervous system. These may finally allow us to map the full process through which humans can exert motor control and thus give us the ability to fully recreate this process artificially for those who may need it. Additionally, new developments in

hardware, batteries, and algorithms will enable richer interactions with the world, through more degrees of freedom and better accuracy. I also envision taking full advantage of sensory feedback to develop assistive technologies that allow realistic interactions while also giving users, especially those missing limbs, a sense of ownership of the device helping them interact with the world.

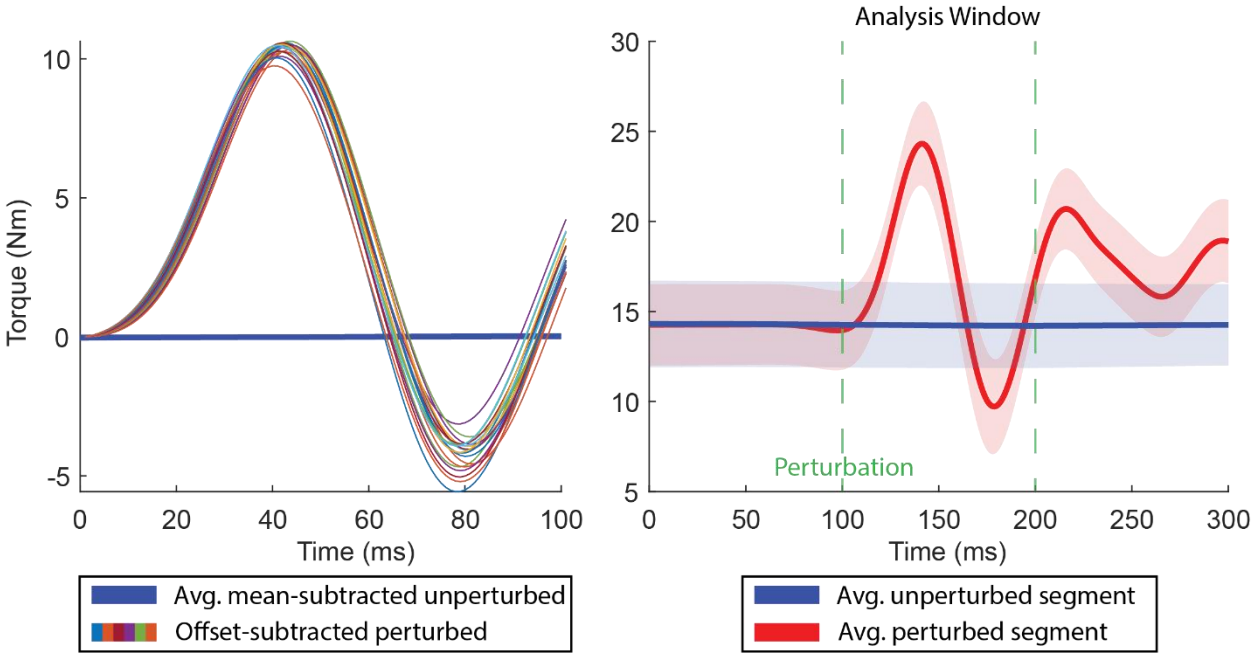
Appendices

Appendix A: Supplement to Chapter 2

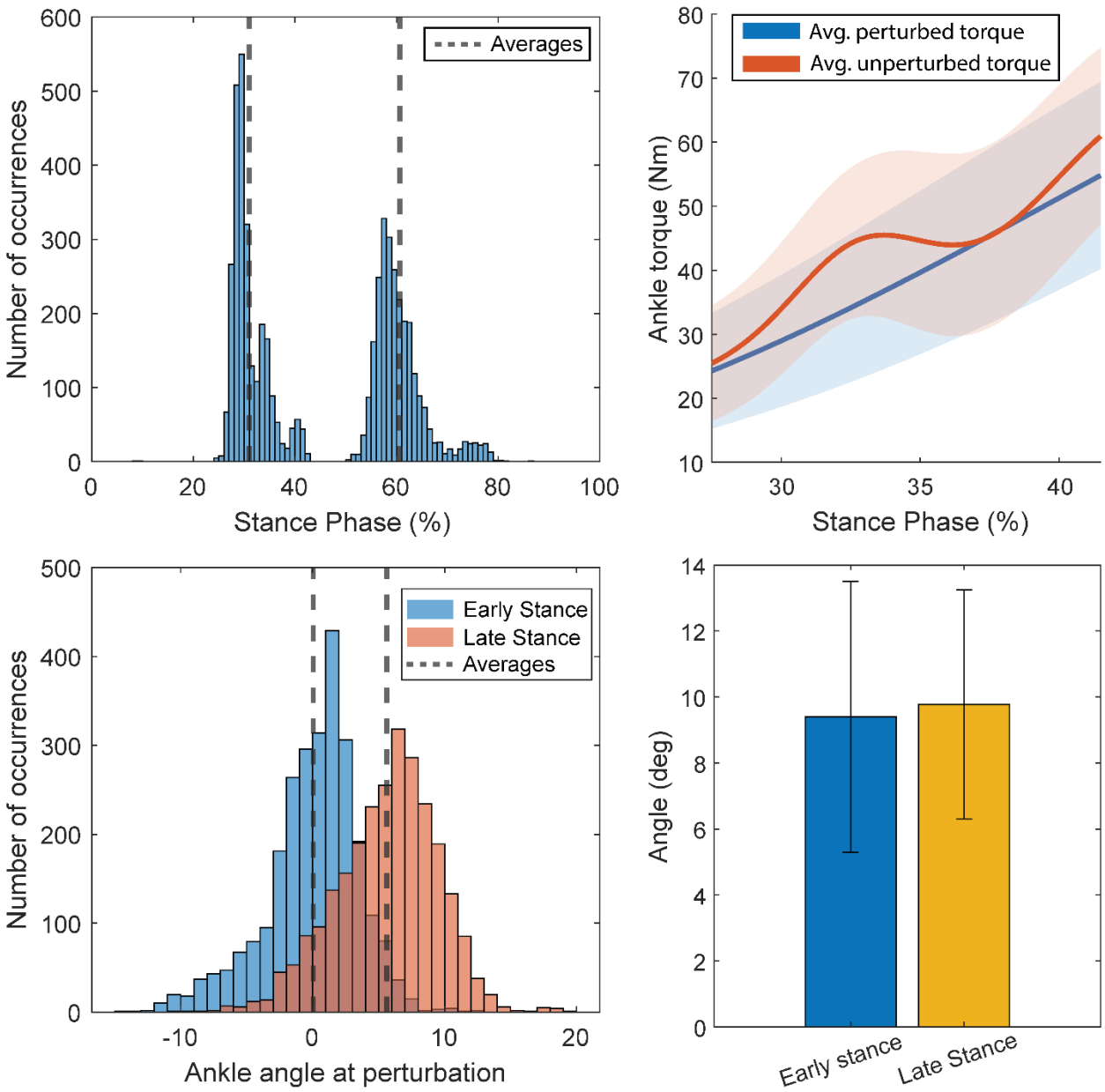


Supp. Figure A.1: **Correlation and Bland-Altman plots for the inertia parameter between sessions for standing and walking**. (A)-(D) show the correlation plots between the Test and Retest sessions for the "only ankle" case (first column) and "ankle + robot" case (second column) during standing (first row) and walking (second row). Blue circles represent values from individual values in the "only ankle" case. Orange squares represent individual values in the "ankle + robot" case. The blue and orange segmented lines represent a linear fit to the data from the "only ankle" and "ankle + robot" cases, respectively, while the black segmented lines show a line of unit slope and zero intercept for reference. The blue and orange text represents the Pearson correlation coefficient between Test and Retest for each corresponding case. (E)-(H) show the Bland-Altman plots for the Test and Retest sessions for the "only ankle" case (first column) and "ankle + robot" case (second column) during standing (first row) and walking (second row). Blue circles and orange squares represent individual values for the "only ankle" and "ankle + robot" cases, respectively. The x-axis on each plot corresponds to the mean between the Test and Retest sessions for each individual value and the y-axis represents the difference between the sessions. The solid blue and orange lines show the mean difference for each case, while the dotted lines represent the 95% confidence interval (1.96 * SD). The

inertia showed good correlation for all but the “only ankle” during walking and was close to the dashed black lines. Bland-Altman plots show good agreement with no apparent systematic errors.



Supp. Figure A.2: **Ankle torque during standing**. Left: ankle torque due to perturbation while standing for a representative subject and session. Each colored trace represents a single perturbed trial, which had its offset (the value right before perturbation) removed. The blue line shows the average of all the 100 ms mean subtracted unperturbed segments before each perturbation. The perturbed torque curves were of greater amplitudes and consistent shapes. Right: comparison between the absolute (non-mean subtracted) ankle torques for 300ms segments of unperturbed trials (blue, shaded region represents standard deviation) and perturbed trials (red, shaded region represents standard deviation). The first green dashed line indicates the moment when perturbation happened and the analysis window started, while the second green dashed line indicates when the analysis window ended.



Supp. Figure A.3: **Perturbation characteristics during walking**. (top, left): Histogram of perturbation timings during walking across subjects and sessions. The vertical black dashed lines represent the average perturbation timing for the early (left line) and late (right line) stance perturbations. (top, right): exemplary ensemble average of early stance perturbed (orange) and unperturbed (blue) ankle torque during walking trials, starting at the average time of perturbation. The perturbed curve shows a more systematic shape, which corresponds to the torque response to the perturbation. (bottom, left): Histogram of ankle angles (in degrees) at perturbation onset during walking, for early (blue) and late (orange) stance. The average ankle angles at each stage are shown with vertical black dashed lines. (bottom, right): Average and standard deviations of the range of ankle angles across all participants for both sessions (range = max. angle in a session – min. angle in a session), separated by when the perturbation happened (early stance: blue, late stance: yellow).

Supp. Table A.1: Descriptive statistics for the impedance parameters while standing and walking for the "ankle+robot" case, separated by sex.

			Test (Mean ± SD)	**Retest (Mean ± SD)**
Standing	**Male**	Stiffness (Nm/rad/kg)	1.35 ± 0.77	1.45 ± 0.92
		Viscosity (Nms/rad/kg)	0.071 ± 0.016	0.071 ± 0.017
		Inertia (m^2/rad)	0.212 ± 0.014	0.213 ± 0.016
	Female	Stiffness (Nm/rad/kg)	1.36 ± 0.63	1.29 ± 0.64
		Viscosity (Nms/rad/kg)	0.058 ± 0.016	0.057 ± 0.013
		Inertia (m^2/rad)	0.212 ± 0.008	0.214 ± 0.017
Walking	**Male**	Stiffness (Nm/rad/kg)	2.51 ± 1.30	2.73 ± 1.39
		Viscosity (Nms/rad/kg)	0.037 ± 0.021	0.039 ± 0.023
		Inertia (m^2/rad)	0.218 ± 0.021	0.222 ± 0.029
	Female	Stiffness (Nm/rad/kg)	2.94 ± 1.39	3.19 ± 0.94
		Viscosity (Nms/rad/kg)	0.061 ± 0.018	0.064 ± 0.022
		Inertia (m^2/rad)	0.200 ± 0.021	0.204 ± 0.025

Supp. Table A.2: Descriptive statistics for the impedance parameters while standing and walking for the "only ankle" case, separated by sex.

			Test (Mean ± SD)	Retest (Mean ± SD)
Standing	**Male**	Stiffness (Nm/rad/kg)	1.52 ± 0.79	1.65 ± 0.95
		Viscosity (Nms/rad/kg)	0.039 ± 0.005	0.041 ± 0.005
		Inertia (m^2/rad)	0.005 ± 0.005	0.005 ± 0.005
	Female	Stiffness (Nm/rad/kg)	1.45 ± 0.51	1.36 ± 0.41
		Viscosity (Nms/rad/kg)	0.036 ± 0.008	0.035 ± 0.005
		Inertia (m^2/rad)	0.009 ± 0.010	0.009 ± 0.008
Walking	**Male**	Stiffness (Nm/rad/kg)	2.12 ± 1.12	2.24 ± 1.08
		Viscosity (Nms/rad/kg)	0.033 ± 0.010	0.037 ± 0.011
		Inertia (m^2/rad)	0.009 ± 0.010	0.008 ± 0.010
	Female	Stiffness (Nm/rad/kg)	2.89 ± 1.14	3.12 ± 0.72
		Viscosity (Nms/rad/kg)	0.031 ± 0.008	0.032 ± 0.007
		Inertia (m^2/rad)	0.007 ± 0.008	0.009 ± 0.008

Supp. Table A.3: Descriptive statistics for the impedance parameters while standing and walking for the "ankle+robot" case, when normalizing by body mass × height. The last column shows the group MDC as a reference for a hypothetical sample size of N=30.

		Test (Mean ± SD)	Retest (Mean ± SD)	ICC [95% CI]	MDC_i	MDC_g (30)
Standing	Stiffness (N/rad/kg)	0.78 ± 0.40	0.78 ± 0.44	0.86 (0.66 - 0.95)	0.430	0.079
	Viscosity (Ns/rad/kg)	0.037 ± 0.011	0.037 ± 0.010	0.93 (0.82 - 0.97)	0.008	0.001
	Inertia (m/rad)	0.123 ± 0.009	0.124 ± 0.012	0.79 (0.53 - 0.92)	0.013	0.002
Walking	Stiffness (N/rad/kg)	1.58 ± 0.81	1.72 ± 0.73	0.75 (0.56 - 0.86)	1.063	0.194
	Viscosity (Ns/rad/kg)	0.028 ± 0.013	0.030 ± 0.016	0.83 (0.70 - 0.91)	0.016	0.003
	Inertia (m/rad)	0.121 ± 0.016	0.123 ± 0.019	0.77 (0.60 - 0.88)	0.024	0.004

Supp. Table A.4: Descriptive statistics for the impedance parameters while standing and walking for the "only ankle" case, when normalizing by body mass × height. The last column shows the group MDC as a reference for a hypothetical sample size of N=30.

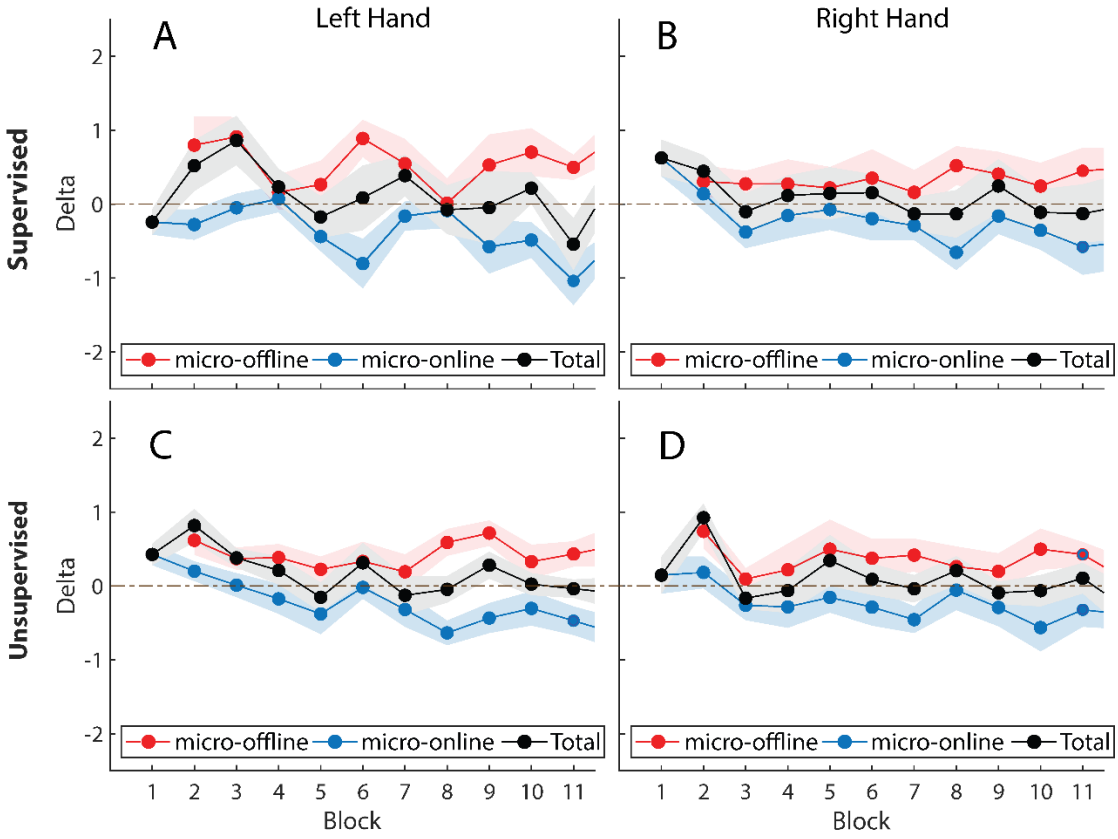
		Test (Mean ± SD)	Retest (Mean ± SD)	ICC [95% CI]	MDC_i	MDC_g (30)
Standing	Stiffness (N/rad/kg)	0.86 ± 0.38	0.86 ± 0.39	0.81 (0.56 - 0.92)	0.460	0.084
	Viscosity (Ns/rad/kg)	0.022 ± 0.004	0.022 ± 0.004	0.75 (0.45 - 0.90)	0.005	0.001
	Inertia (kgm/rad)	0.004 ± 0.005	0.004 ± 0.004	0.85 (0.65 - 0.94)	0.005	0.001
Walking	Stiffness (N/rad/kg)	1.45 ± 0.69	1.55 ± 0.59	0.80 (0.65 - 0.89)	0.798	0.146
	Viscosity (Ns/rad/kg)	0.018 ± 0.005	0.020 ± 0.006	0.77 (0.55 - 0.88)	0.007	0.001
	Inertia (kgm/rad)	0.005 ± 0.005	0.005 ± 0.005	0.44 (0.13 - 0.67)	0.011	0.002

Supp. Table A.5: Descriptive statistics for the absolute (non-normalized) impedance parameters while standing and walking for both analyzed cases.

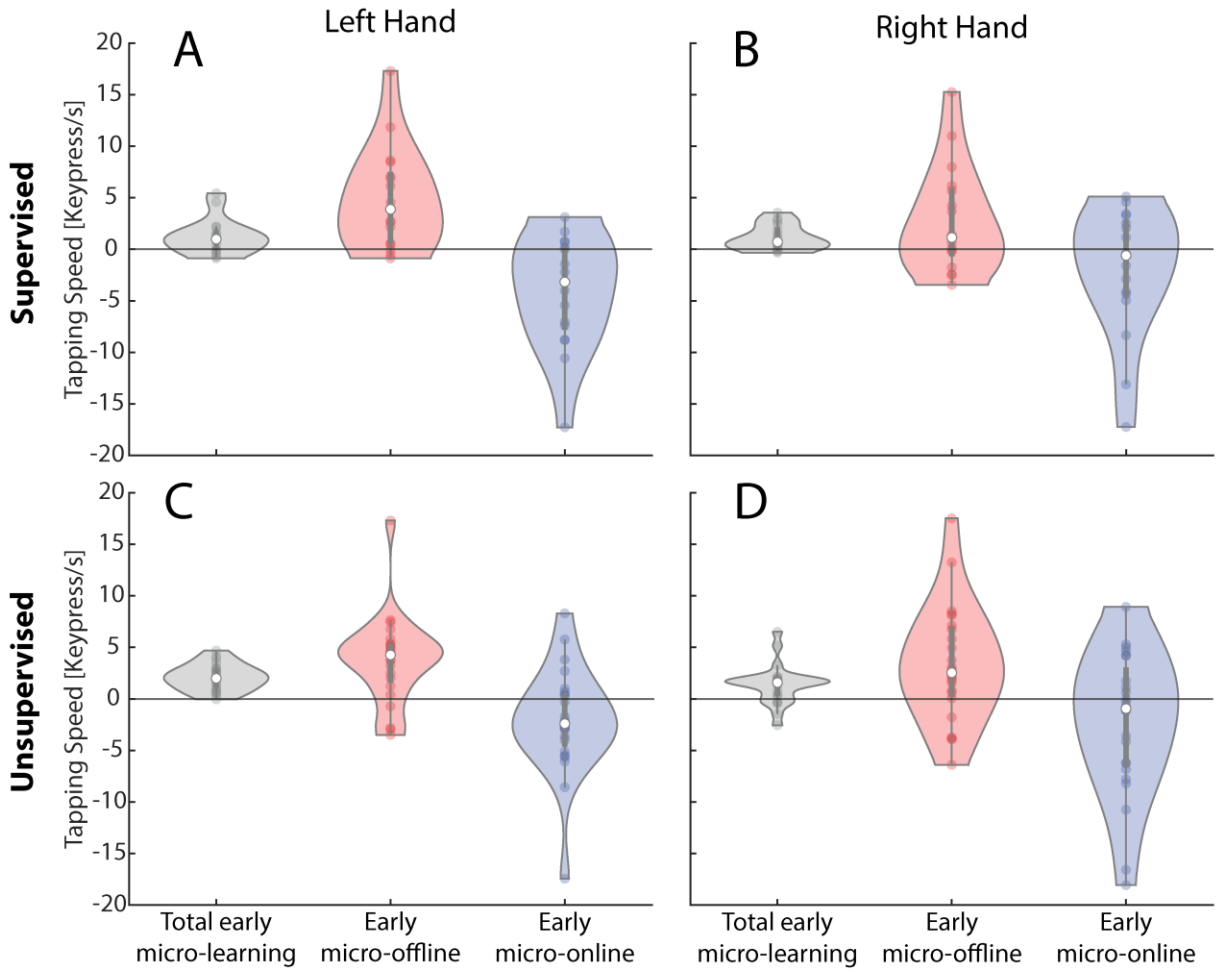
		Ankle + Robot		**Only Ankle**	
		Test (Mean ± SD)	**Retest (Mean ± SD)**	**Test (Mean ± SD)**	**Retest (Mean ± SD)**
Standing	Stiffness (Nm/rad)	99.4 ± 67.4	102.3 ± 81.3	109.1 ± 67.2	111.6 ± 82.6
	Viscosity (Nms/rad)	4.56 ± 1.47	4.54 ± 1.56	2.67 ± 0.72	2.73 ± 0.63
	Inertia (kgm^2/rad)	15.07 ± 2.75	15.16 ± 2.98	0.47 ± 0.61	0.47 ± 0.49
Walking	Stiffness (Nm/rad)	195.6 ± 108.8	211.4 ± 94.9	181.6 ± 100.1	192.4 ± 87.0
	Viscosity (Nms/rad)	3.54 ± 1.83	3.69 ± 1.96	2.28 ± 0.95	2.47 ± 0.99
	Inertia (kgm^2/rad)	14.89 ± 3.52	15.16 ± 3.70	0.58 ± 0.72	0.61 ± 0.72

Appendix B: Supplement to Chapter 3

B.1 Supplementary Figures



Supp. Figure B.1: **Micro-scale learning across groups**: Scatterplot showing the mean micro-scale (micro-offline, micro-online, and total) early learning (first 11 blocks) for the left (A and C), and right (B and D) hand, and for the supervised and unsupervised groups, respectively. Description of the plots is the same as in Figure 3.4



Supp. Figure B.2: **Distribution of micro-scale learning across groups**: violin plots showing the distribution of early micro-offline learning, early micro-online learning, and early total micro-learning for the (A and C) left and (B and D) right hand, and for the supervised and unsupervised groups, respectively. Early micro-learning (micro-offline, micro-online, and total) was computed as the sum of the values observed in the first 11 blocks. Note that the white circles indicate the median values, the filled circles indicate individual data points, and the width of the shaded region represents the approximate frequency of data points in each region.

B.2 Variables Used

A description of all variables used in the study can be found below. T_i represents the average tapping speed on block *i*, $T_{i,1}$ and $T_{i,N}$ represent the tapping speeds on the first and last trial of block *i*, respectively. T_i^L and T_i^R represent the mean tapping speed on block *i* in the left and right hands, respectively. When the superscript is omitted, then it applies to both hands.

1. (Left / right) early micro-online learning: Early micro-online learning is the sum of micro-online learning during the first 11 blocks for each hand. Micro-online learning is defined as the difference in tapping speed between the last and first trial of a practice

block. Micro-online learning quantifies the learning that occurred during short periods of practice.

$$\sum_{i=1}^{11}(T_{i,N} - T_{i,1})$$

2. (Left / right) early micro-offline learning: Early micro-offline learning is the sum of micro-offline learning during the first 11 blocks for each hand. Micro-offline learning is defined as the difference in tapping speed between the first trial of a practice block and the last trial of the previous block. Micro-offline learning quantifies the learning that occurred during short periods of rest.

$$\sum_{i=2}^{11}(T_{i,1} - T_{(i-1),N})$$

3. (Left / right) early total micro-learning: Early total learning is the sum of total micro-learning across the first 11 blocks for each hand and represents all changes in tapping speed during early learning. Total micro-learning is defined as the sum of the micro-offline and micro-online learning for a given block. Total micro-learning quantifies the total learning between blocks (i.e., from the beginning of one block to the beginning of next block).

$$\sum_{i=1}^{11}(T_{i,N} - T_{i,1}) + \sum_{i=2}^{11}(T_{i,1} - T_{(i-1),N})$$

4. Total (left / right) learning: Total learning is defined as the difference between the mean tapping speed of the last and first block for each hand. Total learning quantifies the total increase in tapping speed from the first to the last block of practice.

$$T_{36} - T_1$$

5. Early transfer: Early transfer is the difference between the mean trial tapping speed of the left and right hands on the first block of practice. Early transfer quantifies the amount of improvements in tapping speed on the first block of right hand (relative to the first block of left hand) following the completion of practice with the left hand.

$$T_1^R - T_1^L$$

6. Late transfer: Late transfer is the difference between the mean trial tapping speed of the left and right hands on the last block of practice. Late transfer quantifies the amount of improvements in tapping speed on the last block of right hand (relative to the last block of left hand) following the completion of practice with the left and right hand.

$$T_{36}^R - T_{36}^L$$

B.3 Visualization Plot

For the visualization shown in Figure 4 (C and D), our goal was to show how tapping speed evolved during blocks across all participants. The procedure for computing it was as follows (shown using Matlab notation):

1. We split each 10 seconds block into 10,000 points so that each point represented one millisecond of data. We then denoted the full matrix of data for a single participant *j* as B_j which is of dimensions (1,720000) (10 seconds practice and 10 seconds rest for 36 blocks).
2. For participant *j* and block *i*, we determined how the tapping speed varied across the 10,000 milliseconds. For trial *k* in the corresponding block that had a duration *t* and started at t_0:
    a. If it was a correct trial, we determined the average tapping speed *T* and set it across the duration of the trial

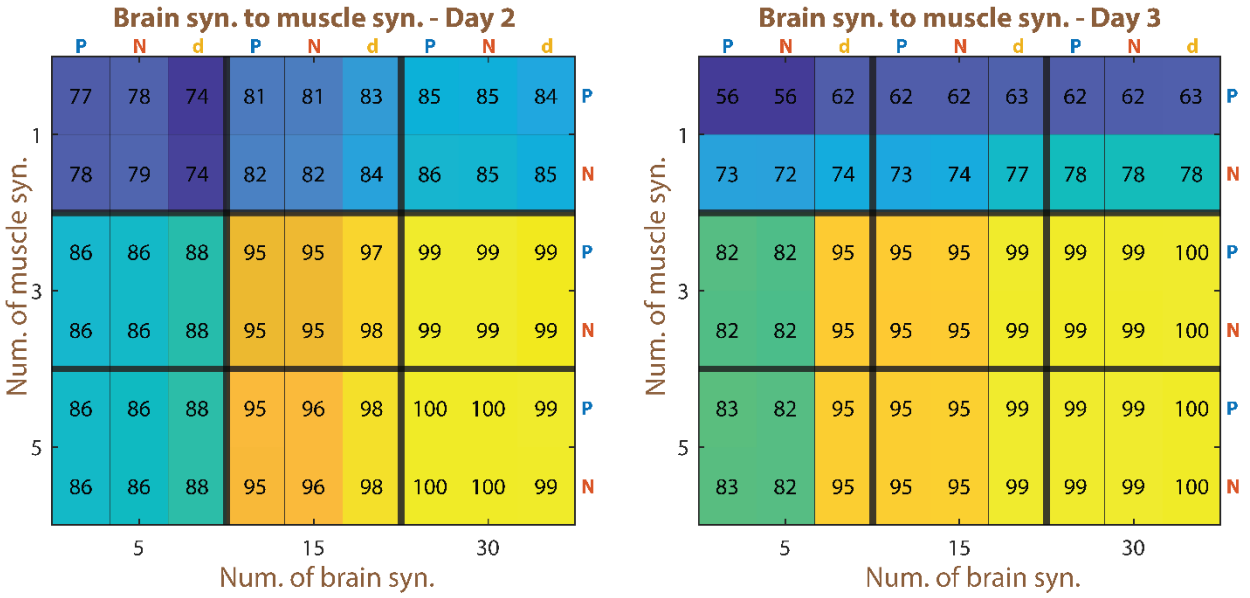
$$B_j(i, t_0 : t_0 + t) = T$$

    b. If it was an incorrect trial, we had two options. If it was the first trial of a block or had no correct trials before it, then we set *T* to be zero across the full duration of the block ($B_j(i, t_0 : t_0 + t) = 0$). If it had a correct trial previously, with tapping speed *T'*, then we set that as the speed for this trial.

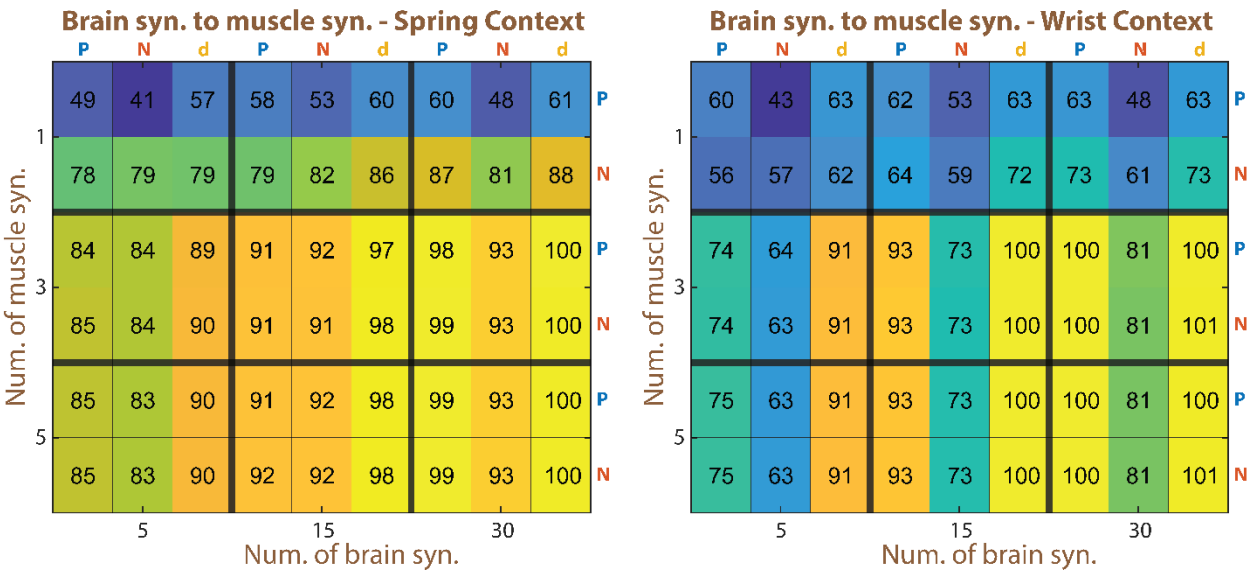
$$B_j(i, t_0 : t_0 + t) = T'$$

3. Finally, we computed an ensemble average across all participants by averaging the tapping speed for each of the 10,000 milliseconds on each block. The resulting matrix was *B*, of dimensions (1,720000), which included the ensemble average across all participants. The shaded region in Figure 4 C and D represents the standard error of the mean.

Appendix C: Supplement to Chapter 4



Supp. Figure C.1: **Heatmap of muscle activity prediction from brain synergies to muscle synergies for day 2** (left) and 3 (right). Each cell value corresponds to the correlation of predicting muscle activity through brain and muscle synergies, normalized to the result of predicting muscle activity directly from brain channels. A value of 100 would mean that it matches the baseline perfectly. P represents PCA, N represents NMF, and d represents dPCA.



Supp. Figure C.2: **Heatmap of muscle activity prediction from brain synergies to muscle synergies the spring** (left) and wrist (right) contexts on day 1. Each cell value corresponds to the correlation of predicting muscle activity through brain and muscle synergies, normalized to the result of the off-context muscle activity prediction directly from brain channels. A value of 100 would mean that it matches the baseline perfectly. P represents PCA, N represents NMF, and d represents dPCA.

Appendix D: Supplement to Chapter 5

D.1 Kalman Filter

Structure

The regular Kalman filter was first introduced by Kalman (Kalman, 1960) and provides a reliable way of tracking state variables over time, given periodic sensor measurements that contain some information about the state. In our case, the state to track is composed of the kinematics of both finger groups:

$$\mathrm{x} = [\mathrm{p_{idx}}\ \mathrm{p_{mrs}}\ \mathrm{v_{idx}}\ \mathrm{v_{mrs}}]$$

Where p_i represents the flexion angle in arbitrary units (from 0 to 1) for finger group i, while v_i represents the flexion velocity in arbitrary units/s for finger group i. The sensor measurements, in our case, correspond to the spiking band power (SBP) of the neural channels. Thus, an observation at time k would be:

$$\mathrm{y_t} = [\mathrm{y_1}\ \mathrm{y_2}\ \ldots\ \mathrm{y_n}]$$

Where n corresponds to the number of neural channels used during the experiment. The KF predicts a new state by first computing the Kalman gain, which uses information about the noise covariance of the dynamics (W) and observations (Q) to determine the optimal interpolation between dynamics and observation predictions:

$$x_{t|t-1} = Ax_{t-1|t-1}$$

$$P_{t|t-1} = AP_{t-1|t-1}A^T + W$$

$$K_t = P_{(t|t-1)}C^T\left[CP_{(t|t-1)}C^T + Q\right]^{(-1)}$$

$$x_{(t|t)} = (I - K_t C)x_{(t|t-1)} + K_t y_t$$

In these equations, A represents the linear dynamical model that encodes the natural evolution of the state through time, C the linear observation model that relates the observations to the state, and $P_{k|k}$ the covariance of the state prediction. Equation 4 uses those models, together with the Kalman gain, to compute a prediction of the state at time t, $x_{t|t}$. When running the Kalman filter online, since the monkey can see in real-time the position of each finger group of the virtual hand, we set the covariance of the position estimate to zero, which has shown to improve results previously (Nason et al., 2021a; Vaskov et al., 2018). Additionally, to control the virtual hand on the screen in online trials, we calculated its position using the following equation: $(\mathrm{pos_t} = 0.02 \cdot \widehat{\mathrm{pos_t}} + 0.98(\mathrm{pos_{t-1}} + 50 \cdot \widehat{\mathrm{vel_t}}))$, which allowed us to interpolate between the position predicted directly ($\widehat{\mathrm{pos}_t}$), and the position predicted through the velocity ($\widehat{\mathrm{vel}_t}$). The velocity term allows for smooth movements, while the position term helps to stabilize the overall position prediction, and the specific values were found empirically.

Model Training

The Kalman filter parameters (dynamics model A, observation model C, and noise covariances Q and W) were computed using a combination of domain knowledge and a data-driven approach. The linear dynamics model A, which represents the evolution of the state through time, had the structure shown in equation 1 (Vaskov et al., 2018). The model integrates velocity (with dt = 50ms being the bin size) and assumes no relationship between positions at time t-1 and those at time t, as well as no relationship between velocities at time t-1 and velocities at time t. The velocity transition sub-matrix, represented in equation 1 with A_{vel}, was learned from an initial calibration run (see Methods) using a least-squares approximation. Note that across days, the diagonal elements of A_{vel} were always strictly greater than 0 and lower than 1, representing

an exponential attenuation of the velocity through time: if no input is applied to the movement, the dynamical model allows for fast stopping.

$$A = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{bmatrix}$$

The observation model C was computed in a similar way. The only difference is that no assumptions were made about the structure, and a bias term was added to account for the difference in base levels. Then, given A and C, we computed the noise covariance for the dynamics (W) and the observations (Q) using the maximum likelihood estimation. For W, we additionally set the noise covariance of the position terms to zero, assuming that the error is only propagated through the velocity term (Vaskov et al., 2018).

In the training of Kalman filter model, an ASUS TUF A15 laptop equipped with an NVIDIA 1660 Ti GPU was utilized. The model took approximately 10 seconds to train for each day.

D.2 KalmanNet

Structure

The main difference between KalmanNet and the Kalman Filter is that KalmanNet uses a set of recurrent neural networks (RNNs) to compute the Kalman Gain. In our implementation, KalmanNet uses the same linear dynamics (A) and observation (C) models, but now does not require an explicit definition of the noise covariances. The full architecture, explained in more detail in (Revach et al., 2022), is shown in Supp. Figure D.1. Briefly, it attempts to compute the Kalman gain by using gated recurrent units (GRUs, a type of recurrent neural network) to implicitly model the noise covariance of the dynamics (GRU 2 in Supp. Figure D.1), the noise covariance of the observations (GRU 1 in Supp. Figure D.1), and the prediction covariance of the

observations (GRU 3 in Supp. Figure D.1). This implicit modeling of the noise covariances allows KalmanNet to dynamically shift the Kalman gain depending on the current state and inputs to choose a better interpolation between states and observations. Overall, the set of equations for KalmanNet is reduced to equations 1 and 4, with K_t corresponding to the output of the network at time t.

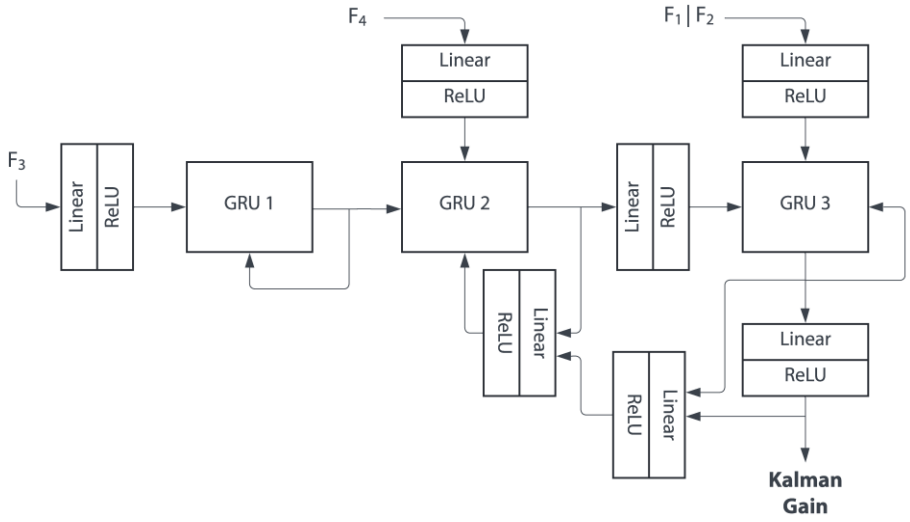
The inputs to the KalmanNet network are variations on the Kalman filter variables: the observation difference (F1), the innovation difference (F2), the forward evolution difference (F3), and the forward update difference (F4) (Revach et al., 2022).

$$\mathrm{F}_1 = \mathrm{y}_\mathrm{t} - \mathrm{y}_{\mathrm{t}-1}$$

$$\mathrm{F}_2 = \mathrm{y}_\mathrm{t} - \mathrm{C}\mathrm{x}_{\mathrm{t}|\mathrm{t}-1}$$

$$\mathrm{F}_3 = \mathrm{x}_{\mathrm{t}-1|\mathrm{t}-1} - x_{t-2|t-2}$$

$$\mathrm{F}_4 = \mathrm{x}_{\mathrm{t}-1|\mathrm{t}-1}$$



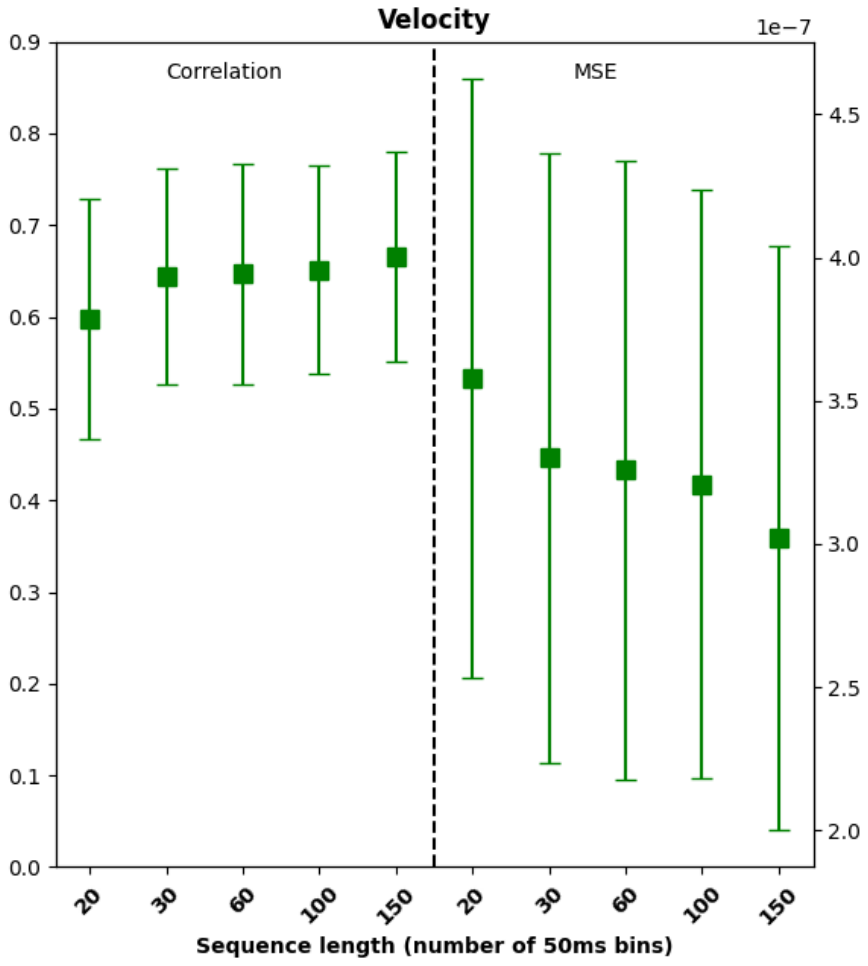
Supp. Figure D.1: **KalmanNet architecture**. Diagram of the components of the KalmanNet network. It consists of three GRUs plus seven linear + ReLU layers that try to model the normal way of computing the Kalman gain (Revach et al., 2022). F1 through F4 correspond to the input features from the previous equations.

When running online, the position of the virtual hand was calculated as for the Kalman filter, integrating the velocity prediction and using 2% of the position prediction for stabilization.

Model Training

To train the KalmanNet network, the loss function used was a scaled mean-squared error (MSE) that accounted for the differences in scale between position and velocity, computed between the predicted and ground truth output. The network does not produce the predicted state directly; instead, it generates the Kalman gain, for which we do not have a ground truth. Consequently, for each time step, a standard update and prediction procedure was executed, incorporating the KalmanNet network output as a component of the computation. Sequences of sixty time bins (equivalent to three seconds of data) were passed through the network, and then the loss was computed as the average scaled MSE between the output at each time step and the ground truth state from the training data.

The KalmanNet network for each day was trained for 300 iterations, each taking a batch of 16 random sequences of length 60 in the training data (Supp. Figure D.2). Since to produce an output, KalmanNet needs to do a forward pass on the network, as well as a Kalman filter step, most of the pyTorch optimizations for network training could not be used, which slowed down training. Overall, the network training took around 10-15 minutes per day on an Asus TUF A15 laptop equipped with an NVIDIA 1660 Ti GPU.



Supp. Figure D.2: **Sensitivity analysis of sequence length during training**. Offline velocity correlation (left) and MSE (right) for KalmanNet, under different sequence lengths employed during training. The horizontal axis represents the number of 50ms bins; the one used throughout corresponds to 60, or equivalently, three seconds. Computed across all n=13 days for both monkeys.

D.3 Heteroscedastic Kalman Filter

The heteroscedastic Kalman filter (HKF) is equivalent to the KF in all but the process noise covariance (W). To enforce a behavior similar to what we observed with KalmanNet's output covarying with the velocity, we forced the dynamics noise covariance in the HKF to also covary with velocity. At time t, and given a base dynamics noise model W, we modified the velocity-related terms using the ground truth velocity of index and MRS at time t, v_t, to produce a new

time-synced noise model W_t (in the following equation). In the new noise model, α corresponds to a constant scaler and v_{max} to the maximum absolute velocity observed across all experiments.

$$W = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{v^{\mathrm{idx}}} & \sigma_{v^{\mathrm{idx}}}\sigma_{v^{\mathrm{mrs}}} \\ 0 & 0 & \sigma_{v^{\mathrm{idx}}}\sigma_{v^{\mathrm{idx}}} & \sigma^2_{v^{\mathrm{mrs}}} \end{bmatrix}$$

$$W_t = \alpha \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{v^{\mathrm{idx}}} \frac{|v_t^{\mathrm{idx}}|}{v_{\max}} & \sigma_{v^{\mathrm{idx}}}\sigma_{v^{\mathrm{mrs}}} \frac{||v_t||}{v_{\max}} \\ 0 & 0 & \sigma_{v^{\mathrm{idx}}}\sigma_{v^{\mathrm{mrs}}} \frac{||v_t||}{v_{\max}} & \sigma^2_{v^{\mathrm{mrs}}} \frac{|v_t^{\mathrm{mrs}}|}{v_{\max}} \end{bmatrix}$$

Note that since we are changing the noise model, rather than the Kalman gain directly, the output Kalman gain of the HKF does not perfectly track the velocity, as the Kalman gain takes a few steps to account for changes in the noise model.

D.4 tcFNN

The tcFNN stands for temporally convolved Feedforward Neural Network, and was introduced by (Willsey et al., 2022) and then further studied by (Temmar et al., 2024). It consists of an initial time convolutional layer that takes in three time bins (150ms) of neural activity followed by four linear layers using ReLU activation functions, batch normalization, and 50% dropout. Note that tcFNN only predicts velocity outputs, rather than position and velocity, as the other decoders tested in this study do. See (Willsey et al., 2022) and (Temmar et al., 2024) for more details on architecture and implementation.

The tcFNN was trained for 15 epochs using the hyperparameters used in (Temmar et al., 2024), and the training took around 1 minute per day on an Asus TUF A15 laptop equipped with an NVIDIA 1660 Ti GPU.

When running online, the position of the virtual hand was calculated by integrating the velocity prediction since no position prediction was available. Additionally, note that previous work with tcFNN has shown that implementing a recalibration feedback intention-training (ReFIT) step after the initial training can substantially improve online performance by assuming that errors during online trials are mostly due to errors in decoder output. In this work, we decided to compare to the base (non-ReFIT) tcFNN, as ReFIT has also been shown to greatly improve performance in the Kalman filter and a similar concept could also be implemented for recurrent approaches, such as the LSTM and the KalmanNet network.

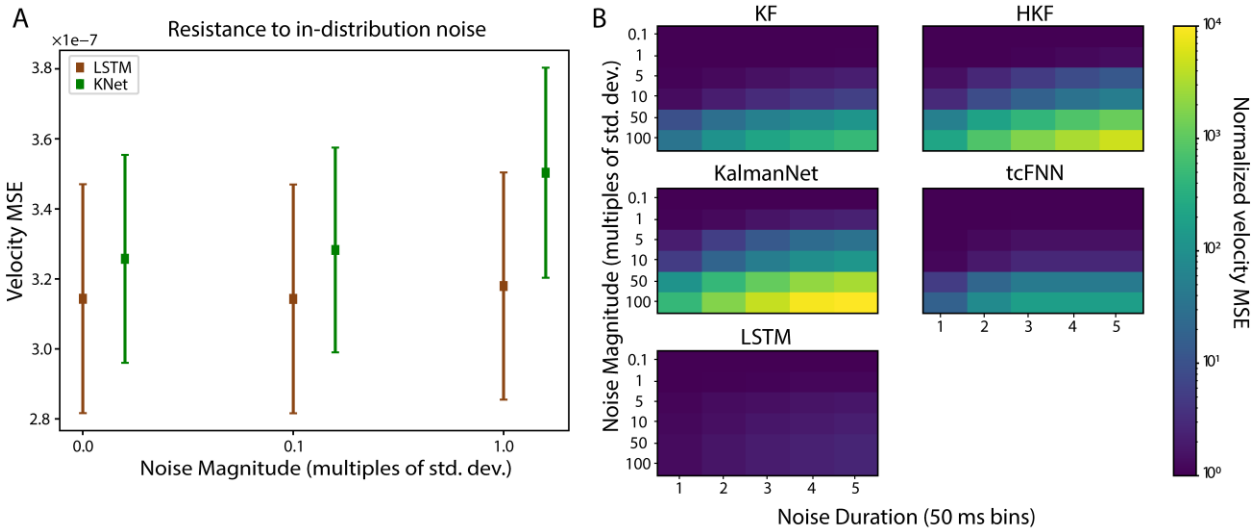
D.5 LSTM

The long short-term memory (LSTM) network is a type of recurrent neural network that has more control over what to remember and what to forget about long-term relationships between inputs. In this study, we refer to LSTM as the architecture used in (Costello et al., 2023), consisting of a single-layer LSTM with a linear layer mapping its hidden output to position and velocity predictions. See (Costello et al., 2023) for more details on architecture and hyperparameters.

The LSTM used the same hyperparameters as in (Costello et al., 2023) and was trained with a scheduler that adapted the learning rate according to the validation loss, and stopped after the validation loss stopped improving. The training took around 2 minutes per day on an Asus TUF A15 laptop equipped with an NVIDIA 1660 Ti GPU.

When running online, the position of the virtual hand was calculated as for KalmanNet and the Kalman filter, integrating the velocity prediction and using 2\% of the position prediction for stabilization.

D.6 Supplemental Figures



Supp. Figure D.3: **Resistance to noise injection**. (A) Offline velocity MSE for KalmanNet (green) and LSTM (brown) across n=13 days for both monkeys, with noise values closer to those present in the training data. A noise of zero magnitude is equivalent to not adding noise (i.e., baseline shown in Figure 5.2). (B) Full product of normalized velocity MSE across models for all values of noise magnitude and duration. The logarithmic color bar on the right represents the MSE value for each combination of noise magnitude and duration, normalized to each model's baseline performance (without noise).

Bibliography

Abbaspourazad, H., Erturk, E., Pesaran, B., & Shanechi, M. M. (2023). Dynamical flexible inference of nonlinear latent factors and structures in neural population activity. *Nature Biomedical Engineering*, 1–24. https://doi.org/10.1038/s41551-023-01106-1

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *WIREs Computational Statistics*, *2*(4), 433–459. https://doi.org/10.1002/wics.101

Adjei, E., Nalam, V., & Lee, H. (2020). Sex Differences in Human Ankle Stiffness During Standing Balance. *Frontiers in Sports and Active Living*, *2*. https://www.frontiersin.org/articles/10.3389/fspor.2020.570449

Aggarwal, C. C. & others. (2018). *Neural networks and deep learning* (Vol. 10). Springer.

Aird, L., Samuel, D., & Stokes, M. (2012). Quadriceps muscle tone, elasticity and stiffness in older males: Reliability and symmetry using the MyotonPRO. *Archives of Gerontology and Geriatrics*, *55*(2), e31–e39. https://doi.org/10.1016/j.archger.2012.03.005

Ajiboye, A. B., & Weir, R. F. (2009). Muscle synergies as a predictive framework for the EMG patterns of new hand postures. *Journal of Neural Engineering*, *6*(3), 036004. https://doi.org/10.1088/1741-2560/6/3/036004

Ajiboye, A. B., Willett, F. R., Young, D. R., Memberg, W. D., Murphy, B. A., Miller, J. P., Walter, B. L., Sweet, J. A., Hoyen, H. A., Keith, M. W., Peckham, P. H., Simeral, J. D., Donoghue, J. P., Hochberg, L. R., & Kirsch, R. F. (2017). Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with

tetraplegia: A proof-of-concept demonstration. *The Lancet*, *389*(10081), 1821–1830. https://doi.org/10.1016/S0140-6736(17)30601-3

Anglada-Tort, M., Harrison, P. M. C., & Jacoby, N. (2022). REPP: A robust cross-platform solution for online sensorimotor synchronization experiments. *Behavior Research Methods*, *54*(5), 2271–2285. https://doi.org/10.3758/s13428-021-01722-2

Ansari, N. N., Naghdi, S., Arab, T. K., & Jalaie, S. (2008). The interrater and intrarater reliability of the Modified Ashworth Scale in the assessment of muscle spasticity: Limb and muscle group effect. *NeuroRehabilitation*, *23*(3), 231–237. https://doi.org/10.3233/NRE-2008-23304

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Armour, B. S., Courtney-Long, E. A., Fox, M. H., Fredine, H., & Cahill, A. (2016). Prevalence and Causes of Paralysis—United States, 2013. *American Journal of Public Health*, *106*(10), 1855–1857. https://doi.org/10.2105/AJPH.2016.303270

Arun, K. S., Huang, T. S., & Blostein, S. D. (1987). Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-9*(5), 698–700. IEEE Transactions on Pattern Analysis and Machine Intelligence. https://doi.org/10.1109/TPAMI.1987.4767965

Au, S., Berniker, M., & Herr, H. (2008). Powered ankle-foot prosthesis to assist level-ground and stair-descent gaits. *Neural Networks*, *21*(4), 654–666. https://doi.org/10.1016/j.neunet.2008.03.006

Azabou, M., Arora, V., Ganesh, V., Mao, X., Nachimuthu, S., Mendelson, M., Richards, B., Perich, M., Lajoie, G., & Dyer, E. (2023). A Unified, Scalable Framework for Neural Population Decoding. *Advances in Neural Information Processing Systems*, *36*, 44937–44956.

Azocar, A. F., Mooney, L. M., Duval, J.-F., Simon, A. M., Hargrove, L. J., & Rouse, E. J. (2020). Design and clinical implementation of an open-source bionic leg. *Nature Biomedical Engineering*, *4*(10), 941–953. https://doi.org/10.1038/s41551-020-00619-3

Balas, M., Roitenberg, N., Giladi, N., & Karni, A. (2007). When practice does not make perfect: Well-practiced handwriting interferes with the consolidation phase gains in learning a movement sequence. *Experimental Brain Research*, *178*(4), 499–508. https://doi.org/10.1007/s00221-006-0757-3

Bark, K., Hyman, E., Tan, F., Cha, E., Jax, S. A., Buxbaum, L. J., & Kuchenbecker, K. J. (2014). Effects of vibrotactile feedback on human learning of arm motions. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *23*(1), 51–63.

Barradas, V. R., Kutch, J. J., Kawase, T., Koike, Y., & Schweighofer, N. (2020). When 90% of the variance is not enough: Residual EMG from muscle synergy extraction influences task performance. *Journal of Neurophysiology*, *123*(6), 2180–2190. https://doi.org/10.1152/jn.00472.2019

Bartko, J. J. (1966). The Intraclass Correlation Coefficient as a Measure of Reliability. *Psychological Reports*, *19*(1), 3–11. https://doi.org/10.2466/pr0.1966.19.1.3

Bear, M., Connors, B., & Paradiso, M. A. (2020). *Neuroscience: Exploring the brain, enhanced edition: exploring the brain*. Jones & Bartlett Learning. https://books.google.com/books?hl=en&lr=&id=m-

PcDwAAQBAJ&oi=fnd&pg=PP1&dq=Neuroscience:+Exploring+the+Brain&ots=Ez9sFaMdsR&sig=ME-aQBTA1lK06w_U4kcMLFg0UWA

Beckerman, H., Roebroeck, M. E., Lankhorst, G. J., Becher, J. G., Bezemer, P. D., & Verbeek, A. L. (2001). Smallest real difference, a link between reproducibility and responsiveness. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *10*(7), 571–578. https://doi.org/10.1023/a:1013138911638

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Best, T. K., Welker, C. G., Rouse, E. J., & Gregg, R. D. (2023). Data-Driven Variable Impedance Control of a Powered Knee–Ankle Prosthesis for Adaptive Speed and Incline Walking. *IEEE Transactions on Robotics*, *39*(3), 2151–2169. IEEE Transactions on Robotics. https://doi.org/10.1109/TRO.2022.3226887

Bohannon, R. W., & Smith, M. B. (1987). Interrater Reliability of a Modified Ashworth Scale of Muscle Spasticity. *Physical Therapy*, *67*(2), 206–207. https://doi.org/10.1093/ptj/67.2.206

Bönstrup, M., Iturrate, I., Hebart, M. N., Censor, N., & Cohen, L. G. (2020). Mechanisms of offline motor learning at a microscale of seconds in large-scale crowdsourced data. *Npj Science of Learning*, *5*(1), 1–10. https://doi.org/10.1038/s41539-020-0066-9

Bönstrup, M., Iturrate, I., Thompson, R., Cruciani, G., Censor, N., & Cohen, L. G. (2019). A Rapid Form of Offline Consolidation in Skill Learning. *Current Biology*, *29*(8), 1346-1351.e4. https://doi.org/10.1016/j.cub.2019.02.049

Brown, J. D., Paek, A., Syed, M., O'Malley, M. K., Shewokis, P. A., Contreras-Vidal, J. L., Davis, A. J., & Gillespie, R. B. (2013). Understanding the role of haptic feedback in a teleoperated/prosthetic grasp and lift task. *2013 World Haptics Conference (WHC)*, 271–276. https://doi.org/10.1109/WHC.2013.6548420

Buch, E. R., Claudino, L., Quentin, R., Bönstrup, M., & Cohen, L. G. (2021). Consolidation of human skill linked to waking hippocampo-neocortical replay. *Cell Reports*, *35*(10). https://doi.org/10.1016/j.celrep.2021.109193

Buchnik, I., Steger, D., Revach, G., van Sloun, R. J., Routtenberg, T., & Shlezinger, N. (2024). Latent-KalmanNet: Learned Kalman Filtering for Tracking from High-Dimensional Signals. *IEEE Transactions on Signal Processsing*, *72*, 352–367. https://doi.org/10.1109/TSP.2023.3344360

Butts, D. A., & Goldman, M. S. (2006). Tuning Curves, Neuronal Variability, and Sensory Coding. *PLOS Biology*, *4*(4), e92. https://doi.org/10.1371/journal.pbio.0040092

Campanini, I., & Merlo, A. (2009). Reliabilty, smallest real difference and concurrent validity of indices computed from GRF components in gait of stroke patients. *Gait & Posture*, *30*(2), 127–131. https://doi.org/10.1016/j.gaitpost.2009.03.011

Card, N. S., Wairagkar, M., Iacobacci, C., Hou, X., Singer-Clark, T., Willett, F. R., Kunz, E. M., Fan, C., Nia, M. V., Deo, D. R., Srinivasan, A., Choi, E. Y., Glasser, M. F., Hochberg, L. R., Henderson, J. M., Shahlaie, K., Stavisky, S. D., & Brandman, D. M. (2024). An Accurate and Rapidly Calibrating Speech Neuroprosthesis. *New England Journal of Medicine*, *391*(7), 609–618. https://doi.org/10.1056/NEJMoa2314132

Carmena, J. M., Lebedev, M. A., Crist, R. E., O'Doherty, J. E., Santucci, D. M., Dimitrov, D. F., Patil, P. G., Henriquez, C. S., & Nicolelis, M. A. L. (2003). Learning to Control a Brain–

Machine Interface for Reaching and Grasping by Primates. *PLOS Biology*, *1*(2), e42. https://doi.org/10.1371/journal.pbio.0000042

Caruso, M., & Harrington, S. (2024). Prevalence of limb loss and limb difference in the United States: Implications for public policy. *White Paper. Avalere and Amputation Coalition*.

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior*, *29*(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Casson, A. J., Yates, D. C., Smith, S. J. M., Duncan, J. S., & Rodriguez-Villegas, E. (2010). Wearable Electroencephalography. *IEEE Engineering in Medicine and Biology Magazine*, *29*(3), 44–56. IEEE Engineering in Medicine and Biology Magazine. https://doi.org/10.1109/MEMB.2010.936545

Censor, N., Sagi, D., & Cohen, L. G. (2012). Common mechanisms of human perceptual and motor learning. *Nature Reviews Neuroscience*, *13*(9), 658–664. https://doi.org/10.1038/nrn3315

Chambaron, S., Ginhac, D., & Perruchet, P. (2008). gSRT-Soft: A generic software application and some methodological guidelines to investigate implicit learning through visual-motor sequential tasks. *Behavior Research Methods*, *40*(2), 493–502. https://doi.org/10.3758/BRM.40.2.493

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, *487*(7405), Article 7405. https://doi.org/10.1038/nature11129

Chvatal, S. A., & Ting, L. H. (2012). Voluntary and reactive recruitment of locomotor muscle synergies during perturbed walking. *Journal of Neuroscience*, *32*(35), 12237–12250.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*, 127–137.

Cimolato, A., & Raspopovic, S. (2024). Closing the sensory feedback loop is necessary for effective neurorehabilitation. *PLOS Biology*, *22*(10), e3002866. https://doi.org/10.1371/journal.pbio.3002866

Cole, J., & Waterman, I. (1995). *Pride and a daily marathon*. MIT press. https://books.google.com/books?hl=en&lr=&id=BA8OBiQYhKMC&oi=fnd&pg=PR9&dq=ian+waterman&ots=Tk_ISzyD7x&sig=nMAoKgfXlLxCcZ4tTKzR9u2d5_0

Collinger, J. L., Boninger, M. L., Bruns, T. M., Curley, K., Wang, W., & Weber, D. J. (2013). Functional priorities, assistive technology, and brain-computer interfaces after spinal cord injury. *Journal of Rehabilitation Research and Development*, *50*(2), 145.

Collinger, J. L., Wodlinger, B., Downey, J. E., Wang, W., Tyler-Kabara, E. C., Weber, D. J., McMorland, A. J., Velliste, M., Boninger, M. L., & Schwartz, A. B. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. *The Lancet*, *381*(9866), 557–564. https://doi.org/10.1016/S0140-6736(12)61816-9

Costa-García, Á., Okajima, S., Yang, N., & Shimoda, S. (2023). Artifact removal from sEMG signals recorded during fully unsupervised daily activities. *DIGITAL HEALTH*, *9*, 20552076231164239. https://doi.org/10.1177/20552076231164239

Costello, J., Temmar, H., Cubillos, L., Mender, M., Wallace, D., Willsey, M., Patil, P., & Chestek, C. (2023). Balancing memorization and generalization in RNNs for high

performance brain-machine Interfaces. *Advances in Neural Information Processing Systems*, *36*, 7462–7474.

Cozza, F., Galdi, P., Serra, A., Pasqua, G., Pavone, L., & Tagliaferri, R. (2020). Dimension Reduction Techniques in a Brain–Computer Interface Application. In A. Esposito, M. Faundez-Zanuy, F. C. Morabito, & E. Pasero (Eds.), *Neural Approaches to Dynamics of Signal Exchanges* (pp. 107–118). Springer. https://doi.org/10.1007/978-981-13-8950-4_11

Cubillos, L. H., Augenstein, T. E., Ranganathan, R., & Krishnan, C. (2023). Breaking the barriers to designing online experiments: A novel open-source platform for supporting procedural skill learning experiments. *Computers in Biology and Medicine*, *154*, 106627. https://doi.org/10.1016/j.compbiomed.2023.106627

Cubillos, L. H., Revach, G., Mender, M. J., Costello, J. T., Temmar, H., Hite, A., Zutshi, D., Wallace, D. M., Ni, X., & Kelberman, M. M. (2024). Exploring the trade-off between deep-learning and explainable models for brain-machine interfaces. *bioRxiv*, 2024–10.

Cubillos, L. H., Rouse, E. J., Augenstein, T. E., Joshi, V., Claflin, E. S., & Krishnan, C. (2024). Reliability and minimal detectable change of stiffness and other mechanical properties of the ankle joint in standing and walking. *Gait & Posture*, *108*, 56–62.

Cunningham, J. P., Nuyujukian, P., Gilja, V., Chestek, C. A., Ryu, S. I., & Shenoy, K. V. (2011). A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces. *Journal of Neurophysiology*, *105*(4), 1932–1949. https://doi.org/10.1152/jn.00503.2010

d'Avella, A., Portone, A., Fernandez, L., & Lacquaniti, F. (2006). Control of fast-reaching movements by muscle synergy combinations. *Journal of Neuroscience*, *26*(30), 7791–7810.

d'Avella, A., Saltiel, P., & Bizzi, E. (2003a). Combinations of muscle synergies in the construction of a natural motor behavior. *Nature Neuroscience*, *6*(3), Article 3. https://doi.org/10.1038/nn1010

d'Avella, A., Saltiel, P., & Bizzi, E. (2003b). Combinations of muscle synergies in the construction of a natural motor behavior. *Nature Neuroscience*, *6*(3), Article 3. https://doi.org/10.1038/nn1010

Damiano, D. L., Quinlivan, J. M., Owen, B. F., Payne, P., Nelson, K. C., & Abel, M. F. (2002). What does the Ashworth scale really measure and are instrumented measures more valid and precise? *Developmental Medicine & Child Neurology*, *44*(2), 112–118. https://doi.org/10.1111/j.1469-8749.2002.tb00296.x

Damon, C., Liutkus, A., Gramfort, A., & Essid, S. (2013). Non-negative matrix factorization for single-channel EEG artifact rejection. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 1177–1181. https://doi.org/10.1109/ICASSP.2013.6637836

Dangi, S., Gowda, S., Héliot, R., & Carmena, J. M. (2011). Adaptive Kalman filtering for closed-loop Brain-Machine Interface systems. *2011 5th International IEEE/EMBS Conference on Neural Engineering*, 609–612. https://doi.org/10.1109/NER.2011.5910622

Davoodi, R., & Loeb, G. E. (2011). MSMS software for VR simulations of neural prostheses and patient training and rehabilitation. *MMVR*, 156–162.

Degenhart, A. D., Bishop, W. E., Oby, E. R., Tyler-Kabara, E. C., Chase, S. M., Batista, A. P., & Yu, B. M. (2020). Stabilization of a brain–computer interface via the alignment of low-dimensional spaces of neural activity. *Nature Biomedical Engineering*, *4*(7), 672–685. https://doi.org/10.1038/s41551-020-0542-9

Deng, L., Li, J., Huang, J.-T., Yao, K., Yu, D., Seide, F., Seltzer, M., Zweig, G., He, X., Williams, J., & others. (2013). Recent advances in deep learning for speech research at Microsoft. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 8604–8608.

Deo, D. R., Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2024a). Brain control of bimanual movement enabled by recurrent neural networks. *Scientific Reports*, *14*(1), 1598. https://doi.org/10.1038/s41598-024-51617-3

Deo, D. R., Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2024b). Brain control of bimanual movement enabled by recurrent neural networks. *Scientific Reports*, *14*(1), 1598. https://doi.org/10.1038/s41598-024-51617-3

Desmurget, M., & Grafton, S. (2000). Forward modeling allows feedback control for fast reaching movements. *Trends in Cognitive Sciences*, *4*(11), 423–431. https://doi.org/10.1016/S1364-6613(00)01537-0

Diekelmann, S., & Born, J. (2007). One memory, two ways to consolidate? *Nature Neuroscience*, *10*(9), 1085–1086. https://doi.org/10.1038/nn0907-1085

Dietterich, T. G. (1998). Approximate Statistical Tests for Comparing Supervised Classification Learning Algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Ding, M., Yang, Z., Hong, W., Zheng, W., Zhou, C., Yin, D., Lin, J., Zou, X., Shao, Z., Yang, H., & others. (2021). Cogview: Mastering text-to-image generation via transformers. *Advances in Neural Information Processing Systems*, *34*, 19822–19835.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.

Epro, G., McCrum, C., Mierau, A., Leyendecker, M., Brüggemann, G.-P., & Karamanidis, K. (2018). Effects of triceps surae muscle strength and tendon stiffness on the reactive dynamic stability and adaptability of older female adults during perturbed walking. *Journal of Applied Physiology*, *124*(6), 1541–1549. https://doi.org/10.1152/japplphysiol.00545.2017

Famié, S., Ammi, M., Bourdin, V., & Amorim, M.-A. (2022). Evidence for an internal model of friction when controlling kinetic energy at impact to slide an object along a surface toward a target. *PLOS ONE*, *17*(2), e0264370. https://doi.org/10.1371/journal.pone.0264370

Flesher, S. N., Collinger, J. L., Foldes, S. T., Weiss, J. M., Downey, J. E., Tyler-Kabara, E. C., Bensmaia, S. J., Schwartz, A. B., Boninger, M. L., & Gaunt, R. A. (2016). Intracortical microstimulation of human somatosensory cortex. *Science Translational Medicine*, *8*(361), 361ra141-361ra141. https://doi.org/10.1126/scitranslmed.aaf8083

Flesher, S. N., Downey, J. E., Weiss, J. M., Hughes, C. L., Herrera, A. J., Tyler-Kabara, E. C., Boninger, M. L., Collinger, J. L., & Gaunt, R. A. (2021). A brain-computer interface that evokes tactile sensations improves robotic arm control. *Science*, *372*(6544), 831–836. https://doi.org/10.1126/science.abd0380

Frère, J., & Hug, F. (2012). Between-subject variability of muscle synergies during a complex motor skill. *Frontiers in Computational Neuroscience*, *6*. https://doi.org/10.3389/fncom.2012.00099

Friedman, J., & Korman, M. (2016). Offline Optimization of the Relative Timing of Movements in a Sequence Is Blocked by Retroactive Behavioral Interference. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00623

Gabert, L., Hood, S., Tran, M., Cempini, M., & Lenzi, T. (2020). A Compact, Lightweight Robotic Ankle-Foot Prosthesis: Featuring a Powered Polycentric Design. *IEEE Robotics & Automation Magazine*, *27*(1), 87–102. IEEE Robotics & Automation Magazine. https://doi.org/10.1109/MRA.2019.2955740

Gallego, J. A., Perich, M. G., Chowdhury, R. H., Solla, S. A., & Miller, L. E. (2020). Long-term stability of cortical population dynamics underlying consistent behavior. *Nature Neuroscience*, *23*(2), Article 2. https://doi.org/10.1038/s41593-019-0555-4

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017a). Neural Manifolds for the Control of Movement. *Neuron*, *94*(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025

Gallego, J. A., Perich, M. G., Miller, L. E., & Solla, S. A. (2017b). Neural Manifolds for the Control of Movement. *Neuron*, *94*(5), 978–984. https://doi.org/10.1016/j.neuron.2017.05.025

Gallego, J. A., Perich, M. G., Naufel, S. N., Ethier, C., Solla, S. A., & Miller, L. E. (2018a). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-06560-z

Gallego, J. A., Perich, M. G., Naufel, S. N., Ethier, C., Solla, S. A., & Miller, L. E. (2018b). Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-06560-z

Gaveau, J., Berret, B., Angelaki, D. E., & Papaxanthis, C. (2016). Direction-dependent arm kinematics reveal optimal integration of gravity cues. *eLife*, *5*, e16394. https://doi.org/10.7554/eLife.16394

George, J. A., Kluger, D. T., Davis, T. S., Wendelken, S. M., Okorokova, E. V., He, Q., Duncan, C. C., Hutchinson, D. T., Thumser, Z. C., Beckler, D. T., Marasco, P. D., Bensmaia, S. J., & Clark, G. A. (2019). Biomimetic sensory feedback through peripheral nerve stimulation improves dexterous use of a bionic hand. *Science Robotics*, *4*(32), eaax2352. https://doi.org/10.1126/scirobotics.aax2352

Ghilardi, M. F., Moisello, C., Silvestri, G., Ghez, C., & Krakauer, J. W. (2009). Learning of a Sequential Motor Skill Comprises Explicit and Implicit Components That Consolidate Differently. *Journal of Neurophysiology*, *101*(5), 2218–2229. https://doi.org/10.1152/jn.01138.2007

Gilja, V., Nuyujukian, P., Chestek, C. A., Cunningham, J. P., Yu, B. M., Fan, J. M., Churchland, M. M., Kaufman, M. T., Kao, J. C., Ryu, S. I., & Shenoy, K. V. (2012). A high-performance neural prosthesis enabled by control algorithm design. *Nature Neuroscience*, *15*(12), Article 12. https://doi.org/10.1038/nn.3265

Glickman, M. E., Rao, S. R., & Schultz, M. R. (2014). False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *Journal of Clinical Epidemiology*, *67*(8), 850–857. https://doi.org/10.1016/j.jclinepi.2014.03.012

Gomi, H., & Kawato, M. (1997). Human arm stiffness and equilibrium-point trajectory during multi-joint movement. *Biological Cybernetics*, *76*(3), 163–171. https://doi.org/10.1007/s004220050329

Gonzalez, M. A., Lee, C., Kang, J., Gillespie, R. B., & Gates, D. H. (2021). Getting a Grip on the Impact of Incidental Feedback From Body-Powered and Myoelectric Prostheses. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *29*, 1905–1912. IEEE Transactions on Neural Systems and Rehabilitation Engineering. https://doi.org/10.1109/TNSRE.2021.3111741

Gonzalez, M. A., Vu, P. P., Vaskov, A. K., Cederna, P. S., Chestek, C. A., & Gates, D. H. (2022). Characterizing sensory thresholds and intensity sensitivity of Regenerative Peripheral Nerve Interfaces: A Case Study. *2022 International Conference on Rehabilitation Robotics (ICORR)*, 1–6. https://doi.org/10.1109/ICORR55369.2022.9896481

Grafton, S. T., Hazeltine, E., & Ivry, R. B. (2002). Motor sequence learning with the nondominant left hand. *Experimental Brain Research*, *146*(3), 369–378. https://doi.org/10.1007/s00221-002-1181-y

Greenspon, C. M., Valle, G., Shelchkova, N. D., Hobbs, T. G., Verbaarschot, C., Callier, T., Berger-Wolf, E. I., Okorokova, E. V., Hutchison, B. C., Dogruoz, E., Sobinov, A. R., Jordan, P. M., Weiss, J. M., Fitzgerald, E. E., Prasad, D., Van Driesche, A., He, Q., Liu, F., Kirsch, R. F., … Bensmaia, S. J. (2024). Evoking stable and precise tactile sensations via multi-electrode intracortical microstimulation of the somatosensory cortex. *Nature Biomedical Engineering*, 1–17. https://doi.org/10.1038/s41551-024-01299-z

Guduru, N. (2006). *Text mining with support vector machines and non-negative matrix factorization algorithms* [PhD Thesis, Citeseer]. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=f652a6df24a3020c5db88872682e2e92ddb054cd

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48.

Hargrove, L. J., Young, A. J., Simon, A. M., Fey, N. P., Lipschutz, R. D., Finucane, S. B., Halsne, E. G., Ingraham, K. A., & Kuiken, T. A. (2015). Intuitive Control of a Powered Prosthetic Leg During Ambulation: A Randomized Clinical Trial. *JAMA*, *313*(22), 2244–2252. https://doi.org/10.1001/jama.2015.4527

Haugh, A. B., Pandyan, A. D., & Johnson, G. R. (2006). A systematic review of the Tardieu Scale for the measurement of spasticity. *Disability and Rehabilitation*, *28*(15), 899–907. https://doi.org/10.1080/09638280500404305

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*(2), 556–573. https://doi.org/10.3758/s13428-019-01283-5

Higgins, S. (1991). Motor Skill Acquisition. *Physical Therapy*, *71*(2), 123–139. https://doi.org/10.1093/ptj/71.2.123

Ho, C. Y., & Bendrups, A. P. (2002). Ankle Reflex Stiffness During Unperceived Perturbation of Standing in Elderly Subjects. *The Journals of Gerontology: Series A*, *57*(9), B344–B350. https://doi.org/10.1093/gerona/57.9.B344

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., Haddadin, S., Liu, J., Cash, S. S., van der Smagt, P., & Donoghue, J. P. (2012). Reach and grasp by

people with tetraplegia using a neurally controlled robotic arm. *Nature*, *485*(7398), 372–375. https://doi.org/10.1038/nature11076

Holdefer, R. N., & Miller, L. E. (2002). Primary motor cortical neurons encode functional muscle synergies. *Experimental Brain Research*, *146*(2), 233–243. https://doi.org/10.1007/s00221-002-1166-x

Hotson, G., McMullen, D. P., Fifer, M. S., Johannes, M. S., Katyal, K. D., Para, M. P., Armiger, R., Anderson, W. S., Thakor, N. V., Wester, B. A., & Crone, N. E. (2016). Individual finger control of a modular prosthetic limb using high-density electrocorticography in a human subject. *Journal of Neural Engineering*, *13*(2), 026017. https://doi.org/10.1088/1741-2560/13/2/026017

Hug, F., Turpin, N. A., Guével, A., & Dorel, S. (2010). Is interindividual variability of EMG patterns in trained cyclists related to different muscle synergies? *Journal of Applied Physiology*, *108*(6), 1727–1736. https://doi.org/10.1152/japplphysiol.01305.2009

Hunter, I. W., & Kearney, R. E. (1982). Dynamics of human ankle stiffness: Variation with mean ankle torque. *Journal of Biomechanics*, *15*(10), 747–752. https://doi.org/10.1016/0021-9290(82)90089-6

Islam, Md. K., Rastegarnia, A., & Sanei, S. (2021). Signal Artifacts and Techniques for Artifacts and Noise Removal. In M. A. R. Ahad & M. U. Ahmed (Eds.), *Signal Processing Techniques for Computational Health Informatics* (pp. 23–79). Springer International Publishing. https://doi.org/10.1007/978-3-030-54932-9_2

Izawa, J., Rane, T., Donchin, O., & Shadmehr, R. (2008). Motor Adaptation as a Process of Reoptimization. *Journal of Neuroscience*, *28*(11), 2883–2891. https://doi.org/10.1523/JNEUROSCI.5359-07.2008

Jain, N. B., Ayers, G. D., Peterson, E. N., Harris, M. B., Morse, L., O'Connor, K. C., & Garshick, E. (2015). Traumatic Spinal Cord Injury in the United States, 1993-2012. *JAMA*, *313*(22), 2236–2243. https://doi.org/10.1001/jama.2015.6250

Japikse, K. C., Negash, S., Howard, J. H., & Howard, D. V. (2003). Intermanual transfer of procedural learning after extended practice of probabilistic sequences. *Experimental Brain Research*, *148*(1), 38–49. https://doi.org/10.1007/s00221-002-1264-9

Jiao, L., & Zhao, J. (2019). A survey on the new generation of deep learning in image processing. *Ieee Access*, *7*, 172231–172263.

Joshi, V., Rouse, E. J., Claflin, E. S., & Krishnan, C. (2022). How Does Ankle Mechanical Stiffness Change as a Function of Muscle Activation in Standing and During the Late Stance of Walking? *IEEE Transactions on Biomedical Engineering*, *69*(3), 1186–1193. IEEE Transactions on Biomedical Engineering. https://doi.org/10.1109/TBME.2021.3117516

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, *82*(1), 35–45. https://doi.org/10.1115/1.3662552

Kami, A., Meyer, G., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1995). Functional MRI evidence for adult motor cortex plasticity during motor skill learning. *Nature*, *377*(6545), 155–158. https://doi.org/10.1038/377155a0

Kandel, E. R., Koester, J. D., Mack, S. H., & Siegelbaum, S. A. (2021). *Principles of Neural Science, Sixth Edition* (6th edition). McGraw Hill / Medical.

Kang, M., & Park, J. (2020). Contragan: Contrastive learning for conditional image generation. *Advances in Neural Information Processing Systems*, *33*, 21357–21369.

Karni, A., Meyer, G., Rey-Hipolito, C., Jezzard, P., Adams, M. M., Turner, R., & Ungerleider, L. G. (1998). The acquisition of skilled motor performance: Fast and slow experience-driven changes in primary motor cortex. *Proceedings of the National Academy of Sciences*, *95*(3), 861–868. https://doi.org/10.1073/pnas.95.3.861

Karpowicz, B. M., Ali, Y. H., Wimalasena, L. N., Sedler, A. R., Keshtkaran, M. R., Bodkin, K., Ma, X., Miller, L. E., & Pandarinath, C. (2022). *Stabilizing brain-computer interfaces through alignment of latent dynamics* (p. 2022.04.06.487388). bioRxiv. https://doi.org/10.1101/2022.04.06.487388

Kearney, R. E., & Hunter, I. W. (1982). Dynamics of human ankle stiffness: Variation with displacement amplitude. *Journal of Biomechanics*, *15*(10), 753–756. https://doi.org/10.1016/0021-9290(82)90090-2

Kesar, T. M., Binder-Macleod, S. A., Hicks, G. E., & Reisman, D. S. (2011). Minimal detectable change for gait variables collected during treadmill walking in individuals post-stroke. *Gait & Posture*, *33*(2), 314–317. https://doi.org/10.1016/j.gaitpost.2010.11.024

Kim, S., & Lockhart, T. (2012). Lower limb control and mobility following exercise training. *Journal of NeuroEngineering and Rehabilitation*, *9*(1), 15. https://doi.org/10.1186/1743-0003-9-15

Kim, S.-P., Simeral, J. D., Hochberg, L. R., Donoghue, J. P., & Black, M. J. (2008). Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia*. *Journal of Neural Engineering*, *5*(4), 455. https://doi.org/10.1088/1741-2560/5/4/010

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., Qi, X.-L., Romo, R., Uchida, N., & Machens, C. K. (2016). Demixed principal component analysis of neural population data. *eLife*, *5*, e10989. https://doi.org/10.7554/eLife.10989

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Korthauer, K., Kimes, P. K., Duvallet, C., Reyes, A., Subramanian, A., Teng, M., Shukla, C., Alm, E. J., & Hicks, S. C. (2019). A practical guide to methods controlling false discoveries in computational biology. *Genome Biology*, *20*(1), 118. https://doi.org/10.1186/s13059-019-1716-1

Korzeczek, A., Cholin, J., Jorschick, A., Hewitt, M., & Sommer, M. (2020). Finger Sequence Learning in Adults Who Stutter. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.01543

Krakauer, J. W. (2006). Motor learning: Its relevance to stroke recovery and neurorehabilitation. *Current Opinion in Neurology*, *19*(1), 84. https://doi.org/10.1097/01.wco.0000200544.29915.cc

Krakauer, J. W., Hadjiosif, A. M., Xu, J., Wong, A. L., & Haith, A. M. (2019). Motor Learning. In *Comprehensive Physiology* (pp. 613–663). John Wiley & Sons, Ltd. https://doi.org/10.1002/cphy.c170043

Krishnan, C., Dharia, A. K., Augenstein, T. E., Washabaugh, E. P., Reid, C. E., Brown, S. R., & Ranganathan, R. (2019). Learning new gait patterns is enhanced by specificity of training rather than progression of task difficulty. *Journal of Biomechanics*, *88*, 33–37. https://doi.org/10.1016/j.jbiomech.2019.03.014

Krishnan, C., Ranganathan, R., & Tetarbe, M. (2017). Interlimb transfer of motor skill learning during walking: No evidence for asymmetric transfer. *Gait & Posture*, *56*, 24–30. https://doi.org/10.1016/j.gaitpost.2017.04.032

Kurtzer, I., Pruszynski, J. A., Herter, T. M., & Scott, S. H. (2006). Primate Upper Limb Muscles Exhibit Activity Patterns That Differ From Their Anatomical Action During a Postural Task. *Journal of Neurophysiology*, *95*(1), 493–504. https://doi.org/10.1152/jn.00706.2005

Kutch, J. J., Kuo, A. D., Bloch, A. M., & Rymer, W. Z. (2008). Endpoint Force Fluctuations Reveal Flexible Rather Than Synergistic Patterns of Muscle Cooperation. *Journal of Neurophysiology*, *100*(5), 2455–2471. https://doi.org/10.1152/jn.90274.2008

Kutch, J. J., & Valero-Cuevas, F. J. (2012a). Challenges and New Approaches to Proving the Existence of Muscle Synergies of Neural Origin. *PLOS Computational Biology*, *8*(5), e1002434. https://doi.org/10.1371/journal.pcbi.1002434

Kutch, J. J., & Valero-Cuevas, F. J. (2012b). Challenges and new approaches to proving the existence of muscle synergies of neural origin. *PLoS Computational Biology*, *8*(5), e1002434.

Lalazar, H., Abbott, L. F., & Vaadia, E. (2016). Tuning Curves for Arm Posture Control in Motor Cortex Are Consistent with Random Connectivity. *PLOS Computational Biology*, *12*(5), e1004910. https://doi.org/10.1371/journal.pcbi.1004910

Lasfargues, J. E., Custis, D., Morrone, F., Cars well, J., & Nguyen, T. (1995). A model for estimating spinal cord injury prevalence in the United States. *Spinal Cord*, *33*(2), 62–68. https://doi.org/10.1038/sc.1995.16

Lavrysen, A., Helsen, W. F., Tremblay, L., Elliott, D., Adam, J. J., Feys, P., & Buekers, M. J. (2003). The Control of Sequential Aiming Movements: The Influence of Practice and Manual Asymmetries On the One-Target Advantage. *Cortex*, *39*(2), 307–325. https://doi.org/10.1016/S0010-9452(08)70111-4

Lee, H., Ho, P., Rastgaar, M., Krebs, H. I., & Hogan, N. (2014). Multivariable Static Ankle Mechanical Impedance With Active Muscles. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *22*(1), 44–52. IEEE Transactions on Neural Systems and Rehabilitation Engineering. https://doi.org/10.1109/TNSRE.2013.2262689

Lee, H., Rouse, E. J., & Krebs, H. I. (2016). Summary of Human Ankle Mechanical Impedance During Walking. *IEEE Journal of Translational Engineering in Health and Medicine*, *4*, 1–7. IEEE Journal of Translational Engineering in Health and Medicine. https://doi.org/10.1109/JTEHM.2016.2601613

Lefumat, H. Z., Vercher, J.-L., Miall, R. C., Cole, J., Buloup, F., Bringoux, L., Bourdin, C., & Sarlegna, F. R. (2015). To transfer or not to transfer? Kinematics and laterality quotient predict interlimb transfer of motor learning. *Journal of Neurophysiology*, *114*(5), 2764–2774. https://doi.org/10.1152/jn.00749.2015

Lenzi, T., Cempini, M., Hargrove, L., & Kuiken, T. (2018). Design, development, and testing of a lightweight hybrid robotic knee prosthesis. *The International Journal of Robotics Research*, *37*(8), 953–976. https://doi.org/10.1177/0278364918785993

Li, F., Wu, Y., & Li, X. (2014). Test-retest reliability and inter-rater reliability of the Modified Tardieu Scale and the Modified Ashworth Scale in hemiplegic patients with stroke. *European Journal of Physical and Rehabilitation Medicine*, *50*(1), 9–15.

Li, Z., O'Doherty, J. E., Hanson, T. L., Lebedev, M. A., Henriquez, C. S., & Nicolelis, M. A. L. (2009). Unscented Kalman Filter for Brain-Machine Interfaces. *PLOS ONE*, *4*(7), e6243. https://doi.org/10.1371/journal.pone.0006243

Lieber, R. L. (2002). *Skeletal muscle structure, function, and plasticity*. Lippincott Williams & Wilkins. https://books.google.com/books?hl=en&lr=&id=T0fbq_b89cAC&oi=fnd&pg=PA1&dq=Skeletal+Muscle+Structure,+Function,+and+Plasticity&ots=jnY6Xa-tRN&sig=LNE6Evssn1Sj9UzznE7JEZd3NNI

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, *23*(1), 18.

Listman, J. B., Tsay, J. S., Kim, H. E., Mackey, W. E., & Heeger, D. J. (2021). Long-Term Motor Learning in the "Wild" With High Volume Video Game Data. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.777779

Lobel, D. A., & Lee, K. H. (2014). Brain Machine Interface and Limb Reanimation Technologies: Restoring Function After Spinal Cord Injury Through Development of a Bypass System. *Mayo Clinic Proceedings*, *89*(5), 708–714. https://doi.org/10.1016/j.mayocp.2014.02.003

Lohse, K., Buchanan, T., & Miller, M. (2016). *Underpowered and Overworked: Problems With Data Analysis in Motor Learning Studies*. https://doi.org/10.1123/jmld.2015-0010

Lu, W.-S., Wang, C.-H., Lin, J.-H., Sheu, C.-F., & Hsieh, C.-L. (2008). The minimal detectable change of the simplified stroke rehabilitation assessment of movement measure. *Journal of Rehabilitation Medicine*, *40*(8), 615–619. https://doi.org/10.2340/16501977-0230

Malik, W. Q., Truccolo, W., Brown, E. N., & Hochberg, L. R. (2011). Efficient Decoding With Steady-State Kalman Filter in Neural Interface Systems. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *19*(1), 25–34. IEEE Transactions on Neural Systems and Rehabilitation Engineering. https://doi.org/10.1109/TNSRE.2010.2092443

Maresova, P., Hruska, J., Klimova, B., Barakovic, S., & Krejcar, O. (2020). Activities of Daily Living and Associated Costs in the Most Widespread Neurodegenerative Diseases: A Systematic Review. *Clinical Interventions in Aging*, *Volume 15*, 1841–1862. https://doi.org/10.2147/CIA.S264688

Marinelli, A., Boccardo, N., Tessari, F., Domenico, D. D., Caserta, G., Canepa, M., Gini, G., Barresi, G., Laffranchi, M., Michieli, L. D., & Semprini, M. (2023). Active upper limb prostheses: A review on current state and upcoming breakthroughs. *Progress in Biomedical Engineering*, *5*(1), 012001. https://doi.org/10.1088/2516-1091/acac57

Markovic, M., Varel, M., Schweisfurth, M. A., Schilling, A. F., & Dosen, S. (2020). Closed-Loop Multi-Amplitude Control for Robust and Dexterous Performance of Myoelectric Prosthesis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(2), 498–507. IEEE Transactions on Neural Systems and Rehabilitation Engineering. https://doi.org/10.1109/TNSRE.2019.2959714

Martin Bland, J., & Altman, DouglasG. (1986). STATISTICAL METHODS FOR ASSESSING AGREEMENT BETWEEN TWO METHODS OF CLINICAL MEASUREMENT. *The Lancet*, *327*(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Mehta, P., Raymond, J., Zhang, Y., Punjani, R., Han, M., Larson, T., Muravov, O., Lyles, R. H., & Horton, D. K. (2023). Prevalence of amyotrophic lateral sclerosis in the United States,

2018. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration*, *24*(7–8), 702–708. https://doi.org/10.1080/21678421.2023.2245858

Mender, M. J., Nason-Tomaszewski, S. R., Temmar, H., Costello, J. T., Wallace, D. M., Willsey, M. S., Ganesh Kumar, N., Kung, T. A., Patil, P., & Chestek, C. A. (2023). The impact of task context on predicting finger movements in a brain-machine interface. *eLife*, *12*, e82598. https://doi.org/10.7554/eLife.82598

Mender, M. J., Ward, A. L., Cubillos, L. H., Kelberman, M. M., Costello, J. T., Temmar, H., Wallace, D. M., Lin, E. T., Lam, J. L. W., Willsey, M. S., Kumar, N. G., Kung, T. A., Patil, P. G., & Chestek, C. A. (2024). *Functional Electrical Stimulation and Brain-Machine Interfaces for Simultaneous Control of Wrist and Finger Flexion* (p. 2024.08.11.607263). bioRxiv. https://doi.org/10.1101/2024.08.11.607263

Meng, X.-L., & Rubin, D. B. (1992). Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika*, *79*(1), 103–111.

Mereu, F., Leone, F., Gentile, C., Cordella, F., Gruppioni, E., & Zollo, L. (2021). Control Strategies and Performance Assessment of Upper-Limb TMR Prostheses: A Review. *Sensors*, *21*(6), Article 6. https://doi.org/10.3390/s21061953

Merletti, R., Aventaggiato, M., Botter, A., Holobar, A., Marateb, H., & Vieira, T. M. (2010). Advances in surface EMG: Recent progress in detection and processing techniques. *Critical Reviews™ in Biomedical Engineering*, *38*(4). https://www.dl.begellhouse.com/journals/4b27cbfc562e21b8,606681d65e2ccb94,45714f720ecf76f5.html

Meseguer-Henarejos, A.-B., Sánchez-Meca, J., López-Pina, J.-A., & Carles-Hernández, R. (2018). Inter- and intra-rater reliability of the Modified Ashworth Scale: A systematic

review and meta-analysis. *European Journal of Physical and Rehabilitation Medicine*, *54*(4), 576–590. https://doi.org/10.23736/s1973-9087.17.04796-7

Metzger, S. L., Littlejohn, K. T., Silva, A. B., Moses, D. A., Seaton, M. P., Wang, R., Dougherty, M. E., Liu, J. R., Wu, P., Berger, M. A., Zhuravleva, I., Tu-Chan, A., Ganguly, K., Anumanchipalli, G. K., & Chang, E. F. (2023). A high-performance neuroprosthesis for speech decoding and avatar control. *Nature*, 1–10. https://doi.org/10.1038/s41586-023-06443-4

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Keith Smith, J. E. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review*, *95*(3), 340–370. https://doi.org/10.1037/0033-295X.95.3.340

Milner, T. E. (1992). A model for the generation of movements requiring endpoint precision. *Neuroscience*, *49*(2), 487–496. https://doi.org/10.1016/0306-4522(92)90113-G

Mistry, M., Mohajerian, P., & Schaal, S. (2005). An exoskeleton robot for human arm movement study. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 4071–4076. https://doi.org/10.1109/IROS.2005.1545450

Mladenić, D. (2006). Feature Selection for Dimensionality Reduction. In C. Saunders, M. Grobelnik, S. Gunn, & J. Shawe-Taylor (Eds.), *Subspace, Latent Structure and Feature Selection* (pp. 84–102). Springer. https://doi.org/10.1007/11752790_5

Moses, D. A., Metzger, S. L., Liu, J. R., Anumanchipalli, G. K., Makin, J. G., Sun, P. F., Chartier, J., Dougherty, M. E., Liu, P. M., Abrams, G. M., & others. (2021). Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *New England Journal of Medicine*, *385*(3), 217–227.

Mougharbel, I., El-Hajj, R., Ghamlouch, H., & Monacelli, E. (2013). Comparative study on different adaptation approaches concerning a sip and puff controller for a powered wheelchair. *2013 Science and Information Conference*, 597–603. https://ieeexplore.ieee.org/abstract/document/6661799

Musk, E., & Neuralink. (2019). An Integrated Brain-Machine Interface Platform With Thousands of Channels. *Journal of Medical Internet Research*, *21*(10), e16194. https://doi.org/10.2196/16194

Nalam, V., Adjei, E., & Lee, H. (2021). Quantification and Modeling of Ankle Stiffness During Standing Balance. *IEEE Transactions on Biomedical Engineering*, *68*(6), 1828–1837. IEEE Transactions on Biomedical Engineering. https://doi.org/10.1109/TBME.2020.3023328

Nalam, V., & Lee, H. (2017). A new robotic approach to characterize mechanical impedance and energetic passivity of the human ankle during standing. *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4123–4126. https://doi.org/10.1109/EMBC.2017.8037763

Nason, S. R., Mender, M. J., Vaskov, A. K., Willsey, M. S., Ganesh Kumar, N., Kung, T. A., Patil, P. G., & Chestek, C. A. (2021a). Real-time linear prediction of simultaneous and independent movements of two finger groups using an intracortical brain-machine interface. *Neuron*, *109*(19), 3164-3177.e8. https://doi.org/10.1016/j.neuron.2021.08.009

Nason, S. R., Mender, M. J., Vaskov, A. K., Willsey, M. S., Ganesh Kumar, N., Kung, T. A., Patil, P. G., & Chestek, C. A. (2021b). Real-time linear prediction of simultaneous and independent movements of two finger groups using an intracortical brain-machine interface. *Neuron*, *109*(19), 3164-3177.e8. https://doi.org/10.1016/j.neuron.2021.08.009

Nason, S. R., Vaskov, A. K., Willsey, M. S., Welle, E. J., An, H., Vu, P. P., Bullard, A. J., Nu, C. S., Kao, J. C., Shenoy, K. V., Jang, T., Kim, H.-S., Blaauw, D., Patil, P. G., & Chestek, C. A. (2020a). A low-power band of neuronal spiking activity dominated by local single units improves the performance of brain–machine interfaces. *Nature Biomedical Engineering*, *4*(10), Article 10. https://doi.org/10.1038/s41551-020-0591-0

Nason, S. R., Vaskov, A. K., Willsey, M. S., Welle, E. J., An, H., Vu, P. P., Bullard, A. J., Nu, C. S., Kao, J. C., Shenoy, K. V., Jang, T., Kim, H.-S., Blaauw, D., Patil, P. G., & Chestek, C. A. (2020b). A low-power band of neuronal spiking activity dominated by local single units improves the performance of brain–machine interfaces. *Nature Biomedical Engineering*, *4*(10), Article 10. https://doi.org/10.1038/s41551-020-0591-0

Newell, K. M. (1991). Motor skill acquisition. *Annual Review of Psychology*, *42*, 213–237. https://doi.org/10.1146/annurev.ps.42.020191.001241

Nguyen, A. T., Drealan, M. W., Luu, D. K., Jiang, M., Xu, J., Cheng, J., Zhao, Q., Keefer, E. W., & Yang, Z. (2021). A portable, self-contained neuroprosthetic hand with deep learning-based finger control. *Journal of Neural Engineering*, *18*(5), 056051.

Ni, X., Revach, G., & Shlezinger, N. (2024). Adaptive KalmanNet: Data-Driven Kalman Filter with Fast Adaptation. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Niegowski, M., Zivanovic, M., Gómez, M., & Lecumberri, P. (2015). Unsupervised learning technique for surface electromyogram denoising from power line interference and baseline wander. *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 7274–7277. https://doi.org/10.1109/EMBC.2015.7320071

Nordhausen, C. T., Rousche, P. J., & Normann, R. A. (1994). Optimizing recording capabilities of the Utah Intracortical Electrode Array. *Brain Research*, *637*(1), 27–36. https://doi.org/10.1016/0006-8993(94)91213-0

Oh, S. J., Schiele, B., & Fritz, M. (2019). Towards reverse-engineering black-box neural networks. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, 121–144.

Okorokova, E. V., He, Q., & Bensmaia, S. J. (2018). Biomimetic encoding model for restoring touch in bionic hands through a nerve interface. *Journal of Neural Engineering*, *15*(6), 066033. https://doi.org/10.1088/1741-2552/aae398

Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*, *32*(2), 604–624.

Overduin, S. A., d'Avella, A., Roh, J., & Bizzi, E. (2008). Modulation of muscle synergy recruitment in primate grasping. *Journal of Neuroscience*, *28*(4), 880–892.

Overduin, S. A., d'Avella, A., Roh, J., Carmena, J. M., & Bizzi, E. (2015a). Representation of Muscle Synergies in the Primate Brain. *Journal of Neuroscience*, *35*(37), 12615–12624. https://doi.org/10.1523/JNEUROSCI.4302-14.2015

Overduin, S. A., d'Avella, A., Roh, J., Carmena, J. M., & Bizzi, E. (2015b). Representation of Muscle Synergies in the Primate Brain. *Journal of Neuroscience*, *35*(37), 12615–12624. https://doi.org/10.1523/JNEUROSCI.4302-14.2015

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., Henderson, J. M., Shenoy, K. V., Abbott, L. F., & Sussillo, D. (2018). Inferring single-trial neural

population dynamics using sequential auto-encoders. *Nature Methods*, *15*(10), Article 10. https://doi.org/10.1038/s41592-018-0109-9

Paolacci, G., & Chandler, J. (2014). Inside the Turk: Understanding Mechanical Turk as a Participant Pool. *Current Directions in Psychological Science*, *23*(3), 184–188. https://doi.org/10.1177/0963721414531598

Parlow, S. E., & Kinsbourne, M. (1989). Asymmetrical transfer of training between hands: Implications for interhemispheric communication in normal brain. *Brain and Cognition*, *11*(1), 98–113. https://doi.org/10.1016/0278-2626(89)90008-0

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Perez, M. A., Wise, S. P., Willingham, D. T., & Cohen, L. G. (2007). Neurophysiological mechanisms involved in transfer of procedural knowledge. *Journal of Neuroscience*, *27*(5), 1045–1053. Scopus. https://doi.org/10.1523/JNEUROSCI.4128-06.2007

Perge, J. A., Homer, M. L., Malik, W. Q., Cash, S., Eskandar, E., Friehs, G., Donoghue, J. P., & Hochberg, L. R. (2013). Intra-day signal instabilities affect decoding performance in an intracortical neural interface system. *Journal of Neural Engineering*, *10*(3), 036004. https://doi.org/10.1088/1741-2560/10/3/036004

Perreault, E. J., Kirsch, R. F., & Acosta, A. M. (1999). Multiple-input, multiple-output system identification for characterization of limb stiffness dynamics. *Biological Cybernetics*, *80*(5), 327–337. https://doi.org/10.1007/s004220050529

Plamondon, R., & Alimi, A. M. (1997). Speed/accuracy trade-offs in target-directed movements. *Behavioral and Brain Sciences*, *20*(2), 279–303.

Popescu, F., Hidler, J. M., & Rymer, W. Z. (2003). Elbow impedance during goal-directed movements. *Experimental Brain Research*, *152*(1), 17–28. https://doi.org/10.1007/s00221-003-1507-4

Prechtl, H. F., Cioni, G., Einspieler, C., Bos, A. F., & Ferrari, F. (2001). Role of vision on early motor development: Lessons from the blind. *Developmental Medicine and Child Neurology*, *43*(3), 198–201. https://doi.org/10.1017/S0012162201000378

Rabbi, M. F., Pizzolato, C., Lloyd, D. G., Carty, C. P., Devaprakash, D., & Diamond, L. E. (2020). Non-negative matrix factorisation is the most appropriate method for extraction of muscle synergies in walking and running. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-65257-w

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Ranganathan, R., & Krishnan, C. (2012). Extracting synergies in gait: Using EMG variability to evaluate control strategies. *Journal of Neurophysiology*, *108*(5), 1537–1544. https://doi.org/10.1152/jn.01112.2011

Ranganathan, R., Krishnan, C., Dhaher, Y. Y., & Rymer, W. Z. (2016). Learning new gait patterns: Exploratory muscle activity during motor learning is not predicted by motor modules. *Journal of Biomechanics*, *49*(5), 718–725.

Ranganathan, R., LEE, M.-H., & KRISHNAN, C. (2022). Ten guidelines for designing motor learning studies. *Brazilian Journal of Motor Behavior*, *16*(2), 112–133. https://doi.org/10.20338/bjmb.v16i2.283

Ranganathan, R., Tomlinson, A. D., Lokesh, R., Lin, T.-H., & Patel, P. (2021). A tale of too many tasks: Task fragmentation in motor learning and a call for model task paradigms. *Experimental Brain Research*, *239*(1), 1–19. https://doi.org/10.1007/s00221-020-05908-6

Ranjan, R., Sankaranarayanan, S., Bansal, Ankan, Bodla, N., Chen, J.-C., Patel, V. M., Castillo, C. D., & Chellappa, R. (2018). Deep learning for understanding faces: Machines may be just as good, or better, than humans. *IEEE Signal Processing Magazine*, *35*(1), 66–83.

Revach, G., Shlezinger, N., Ni, X., Escoriza, A. L., van Sloun, R. J. G., & Eldar, Y. C. (2022). KalmanNet: Neural Network Aided Kalman Filtering for Partially Known Dynamics. *IEEE Transactions on Signal Processing*, *70*, 1532–1547. IEEE Transactions on Signal Processing. https://doi.org/10.1109/TSP.2022.3158588

Robertson, E. M. (2007). The Serial Reaction Time Task: Implicit Motor Skill Learning? *Journal of Neuroscience*, *27*(38), 10073–10075. https://doi.org/10.1523/JNEUROSCI.2747-07.2007

Robertson, E. M. (2019). Skill Memory: Mind the Ever-Decreasing Gap for Offline Processing. *Current Biology*, *29*(8), R287–R289. https://doi.org/10.1016/j.cub.2019.03.007

Rohde, M., & Ernst, M. O. (2016). Time, agency, and sensory feedback delays during action. *Current Opinion in Behavioral Sciences*, *8*, 193–199. https://doi.org/10.1016/j.cobeha.2016.02.029

Rouse, E. J., Gregg, R. D., Hargrove, L. J., & Sensinger, J. W. (2013). The difference between stiffness and quasi-stiffness in the context of biomechanical modeling. *IEEE Transactions on Bio-Medical Engineering*, *60*(2), 562–568. https://doi.org/10.1109/TBME.2012.2230261

Rouse, E. J., Hargrove, L. J., Perreault, E. J., & Kuiken, T. A. (2014). Estimation of Human Ankle Impedance During the Stance Phase of Walking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *22*(4), 870–878. IEEE Transactions on Neural Systems and Rehabilitation Engineering. https://doi.org/10.1109/TNSRE.2014.2307256

Rouse, E. J., Hargrove, L. J., Perreault, E. J., Peshkin, M. A., & Kuiken, T. A. (2013). Development of a Mechatronic Platform and Validation of Methods for Estimating Ankle Stiffness During the Stance Phase of Walking. *Journal of Biomechanical Engineering*, *135*(081009). https://doi.org/10.1115/1.4024286

Roy, A., Krebs, H. I., Bever, C. T., Forrester, L. W., Macko, R. F., & Hogan, N. (2011). Measurement of passive ankle stiffness in subjects with chronic hemiparesis using a novel ankle robot. *Journal of Neurophysiology*, *105*(5), 2132–2149. https://doi.org/10.1152/jn.01014.2010

Russo, A. A., Bittner, S. R., Perkins, S. M., Seely, J. S., London, B. M., Lara, A. H., Miri, A., Marshall, N. J., Kohn, A., Jessell, T. M., Abbott, L. F., Cunningham, J. P., & Churchland, M. M. (2018). Motor Cortex Embeds Muscle-like Commands in an Untangled Population Response. *Neuron*, *97*(4), 953-966.e8. https://doi.org/10.1016/j.neuron.2018.01.004

Safdarian, M., Trinka, E., Rahimi-Movaghar, V., Thomschewski, A., Aali, A., Abady, G. G., Abate, S. M., Abd-Allah, F., Abedi, A., Adane, D. E., Afzal, S., Ahinkorah, B. O., Ahmad, S., Ahmed, H., Amanat, N., Angappan, D., Arabloo, J., Aryannejad, A., Athari, S. S., … Zhang, Z.-J. (2023). Global, regional, and national burden of spinal cord injury, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet Neurology*, *22*(11), 1026–1047. https://doi.org/10.1016/S1474-4422(23)00287-9

Samek, W., Wiegand, T., & Müller, K.-R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv Preprint arXiv:1708.08296*.

Sani, O. G., Abbaspourazad, H., Wong, Y. T., Pesaran, B., & Shanechi, M. M. (2021). Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification. *Nature Neuroscience*, *24*(1), Article 1. https://doi.org/10.1038/s41593-020-00733-0

Saunders, I., & Vijayakumar, S. (2011). The role of feed-forward and feedback processes for closed-loop prosthesis control. *Journal of NeuroEngineering and Rehabilitation*, *8*(1), 60. https://doi.org/10.1186/1743-0003-8-60

Saxena, S., Russo, A. A., Cunningham, J., & Churchland, M. M. (2022). Motor cortex activity across movement speeds is predicted by network-level strategies for generating muscle activity. *eLife*, *11*, e67620. https://doi.org/10.7554/eLife.67620

Scheid, F. (1968). *Schaum's Outline Series Theory and Problems of Numerical Analysis*. McGraw-Hill.

Schmidt, R. A., Lee, T. D., Winstein, C., Wulf, G., & Zelaznik, H. N. (2018). *Motor control and learning: A behavioral emphasis*. Human kinetics. https://books.google.com/books?hl=en&lr=&id=EvJ6DwAAQBAJ&oi=fnd&pg=PR1&dq=Motor+control+and+learning:+A+behavioral+emphasis&ots=k6DndKqaEB&sig=kugOqfi6lrA5gip3Cnfs9i0OzL8

Selen, L. P. J., van Dieën, J. H., & Beek, P. J. (2006). Impedance Modulation and Feedback Corrections in Tracking Targets of Variable Size and Frequency. *Journal of Neurophysiology*, *96*(5), 2750–2759. https://doi.org/10.1152/jn.00552.2006

Semmelmann, K., & Weigelt, S. (2017). Online psychophysics: Reaction time effects in cognitive experiments. *Behavior Research Methods*, *49*(4), 1241–1260. https://doi.org/10.3758/s13428-016-0783-4

Shah, N. P., Willsey, M. S., Hahn, N., Kamdar, F., Avansino, D. T., Fan, C., Hochberg, L. R., Willett, F. R., & Henderson, J. M. (2024). A flexible intracortical brain-computer interface for typing using finger movements. *bioRxiv*, 2024.04.22.590630. https://doi.org/10.1101/2024.04.22.590630

Shlezinger, N., Whang, J., Eldar, Y. C., & Dimakis, A. G. (2023). Model-Based Deep Learning. *Proceedings of the IEEE*, *111*(5), 465–499. Proceedings of the IEEE. https://doi.org/10.1109/JPROC.2023.3247480

Shorter, A. L., Richardson, J. K., Finucane, S. B., Joshi, V., Gordon, K., & Rouse, E. J. (2021). Characterization and clinical implications of ankle impedance during walking in chronic stroke. *Scientific Reports*, *11*(1), Article 1. https://doi.org/10.1038/s41598-021-95737-6

Shorter, A. L., & Rouse, E. J. (2018). Mechanical Impedance of the Ankle During the Terminal Stance Phase of Walking. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *26*(1), 135–143. IEEE Transactions on Neural Systems and Rehabilitation Engineering. https://doi.org/10.1109/TNSRE.2017.2758325

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, *15*(1), 1929–1958.

Steele, K. M., Tresch, M. C., & Perreault, E. J. (2015). Consequences of biomechanically constrained tasks in the design and interpretation of synergy analyses. *Journal of Neurophysiology*, *113*(7), 2102–2113. https://doi.org/10.1152/jn.00769.2013

Stepan, G. (2009). Delay effects in the human sensory system during balancing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1891), 1195–1212. https://doi.org/10.1098/rsta.2008.0278

Stevenson, I. H., & Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience*, *14*(2), 139–142. https://doi.org/10.1038/nn.2731

Stoet, G. (2017). PsyToolkit: A Novel Web-Based Method for Running Online Questionnaires and Reaction-Time Experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Sussillo, D., Nuyujukian, P., Fan, J. M., Kao, J. C., Stavisky, S. D., Ryu, S., & Shenoy, K. (2012). A recurrent neural network for closed-loop intracortical brain–machine interface decoders. *Journal of Neural Engineering*, *9*(2), 026027. https://doi.org/10.1088/1741-2560/9/2/026027

Takei, T., Confais, J., Tomatsu, S., Oya, T., & Seki, K. (2017). Neural basis for hand muscle synergies in the primate spinal cord. *Proceedings of the National Academy of Sciences*, *114*(32), 8643–8648. https://doi.org/10.1073/pnas.1704328114

Tang, Y., Chen, D., & Li, X. (2021). Dimensionality Reduction Methods for Brain Imaging Data Analysis. *ACM Comput. Surv.*, *54*(4), 87:1-87:36. https://doi.org/10.1145/3448302

Temmar, H., Willsey, M. S., Costello, J. T., Mender, M. J., Cubillos, L. H., Lam, J. L., Wallace, D. M., Kelberman, M. M., Patil, P. G., & Chestek, C. A. (2024). *Artificial neural network for brain-machine interface consistently produces more naturalistic finger movements than linear methods* (p. 2024.03.01.583000). bioRxiv. https://doi.org/10.1101/2024.03.01.583000

Thies, J., & Alimohammad, A. (2019). Compact and Low-Power Neural Spike Compression Using Undercomplete Autoencoders. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *27*(8), 1529–1538. IEEE Transactions on Neural Systems and Rehabilitation Engineering. https://doi.org/10.1109/TNSRE.2019.2929081

Tresch, M. C., & Jarc, A. (2009). The case for and against muscle synergies. *Current Opinion in Neurobiology*, *19*(6), 601–607. https://doi.org/10.1016/j.conb.2009.09.002

Truzman, S., Revach, G., Shlezinger, N., & Klein, I. (2023). Outlier-Insensitive Kalman Filtering Using NUV Priors. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Tsay, J. S., Lee, A., Ivry, R. B., & Avraham, G. (2021). Moving outside the lab: The viability of conducting sensorimotor learning studies online. *Neurons, Behavior, Data Analysis, and Theory*, *5*(3), 1–22. https://doi.org/10.51628/001c.26985

Tsui, C. S. L., Gan, J. Q., & Roberts, S. J. (2009). A self-paced brain–computer interface for controlling a robot simulator: An online event labelling paradigm and an extended Kalman filter based algorithm for online training. *Medical & Biological Engineering & Computing*, *47*(3), 257–265. https://doi.org/10.1007/s11517-009-0459-7

Tsuji, T., Takeda, Y., & Tanaka, Y. (2004). Analysis of mechanical impedance in human arm movements using a virtual tennis system. *Biological Cybernetics*, *91*(5), 295–305. https://doi.org/10.1007/s00422-004-0515-1

Tukey, J. W. (1977). Exploratory data analysis. *Reading/Addison-Wesley*.

Turpin, N. A., Uriac, S., & Dalleau, G. (2021). How to improve the muscle synergy analysis methodology? *European Journal of Applied Physiology*, *121*(4), 1009–1025. https://doi.org/10.1007/s00421-021-04604-9

Valero-Cuevas, F. J., Venkadesan, M., & Todorov, E. (2009). Structured Variability of Muscle Activations Supports the Minimal Intervention Principle of Motor Control. *Journal of Neurophysiology*, *102*(1), 59–68. https://doi.org/10.1152/jn.90324.2008

Van Der Werf, Y. D., Van Der Helm, E., Schoonheim, M. M., Ridderikhoff, A., & Van Someren, E. J. W. (2009). Learning by observation requires an early sleep window. *Proceedings of the National Academy of Sciences*, *106*(45), 18926–18930. https://doi.org/10.1073/pnas.0901320106

van Wijck, F. M. J., Pandyan, A. D., Johnson, G. R., & Barnes, M. P. (2001). Assessing Motor Deficits in Neurological Rehabilitation: Patterns of Instrument Usage. *Neurorehabilitation and Neural Repair*, *15*(1), 23–30. https://doi.org/10.1177/154596830101500104

Vaskov, A. K., Irwin, Z. T., Nason, S. R., Vu, P. P., Nu, C. S., Bullard, A. J., Hill, M., North, N., Patil, P. G., & Chestek, C. A. (2018). Cortical Decoding of Individual Finger Group Motions Using ReFIT Kalman Filter. *Frontiers in Neuroscience*, *12*. https://www.frontiersin.org/articles/10.3389/fnins.2018.00751

Vet, H. C. W. de, Bouter, L. M., Bezemer, P. D., & Beurskens, A. J. H. M. (2001). REPRODUCIBILITY AND RESPONSIVENESS OF EVALUATIVE OUTCOME MEASURES: Theoretical Considerations Illustrated by an Empirical Example. *International Journal of Technology Assessment in Health Care*, *17*(4), 479–487. https://doi.org/10.1017/S0266462301107038

Vu, P. P., Lu, C., Vaskov, A. K., Gates, D. H., Brent Gillespie, R., Kung, T. A., Chestek, C. A., Cederna, P. S., & Kemp, S. W. P. (2020). Abstract 100: Restoration Of Proprioceptive And Cutaneous Sensation Using Regenerative Peripheral Nerve Interfaces (RPNIs) In

Humans With Upper-limb Amputations. *Plastic and Reconstructive Surgery – Global Open*, *8*(4S), 65. https://doi.org/10.1097/01.GOX.0000667456.98464.10

Wang, J., & Sainburg, R. L. (2004). Interlimb Transfer of Novel Inertial Dynamics Is Asymmetrical. *Journal of Neurophysiology*, *92*(1), 349–360. https://doi.org/10.1152/jn.00960.2003

Ware, C., & Mikaelian, H. H. (1986). An evaluation of an eye tracker as a device for computer input2. *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, 183–188. https://doi.org/10.1145/29933.275627

Washabaugh, E. P., Kalyanaraman, T., Adamczyk, P. G., Claflin, E. S., & Krishnan, C. (2017). Validity and repeatability of inertial measurement units for measuring gait parameters. *Gait & Posture*, *55*, 87–93. https://doi.org/10.1016/j.gaitpost.2017.04.013

Wei, K., & Körding, K. (2009). Relevance of Error: What Drives Motor Adaptation? *Journal of Neurophysiology*, *101*(2), 655–664. https://doi.org/10.1152/jn.90545.2008

Westling, G., & Johansson, R. S. (1987). Responses in glabrous skin mechanoreceptors during precision grip in humans. *Experimental Brain Research*, *66*(1), 128–140. https://doi.org/10.1007/BF00236209

Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2021). High-performance brain-to-text communication via handwriting. *Nature*, *593*(7858), Article 7858. https://doi.org/10.1038/s41586-021-03506-2

Willett, F. R., Kunz, E. M., Fan, C., Avansino, D. T., Wilson, G. H., Choi, E. Y., Kamdar, F., Glasser, M. F., Hochberg, L. R., Druckmann, S., Shenoy, K. V., & Henderson, J. M. (2023). A high-performance speech neuroprosthesis. *Nature*, *620*(7976), Article 7976. https://doi.org/10.1038/s41586-023-06377-x

Willsey, M. S., Nason-Tomaszewski, S. R., Ensel, S. R., Temmar, H., Mender, M. J., Costello, J. T., Patil, P. G., & Chestek, C. A. (2022). Real-time brain-machine interface in non-human primates achieves high-velocity prosthetic finger movements using a shallow feedforward neural network decoder. *Nature Communications*, *13*(1), Article 1. https://doi.org/10.1038/s41467-022-34452-w

Willsey, M. S., Shah, N. P., Avansino, D. T., Hahn, N. V., Jamiolkowski, R. M., Kamdar, F. B., Hochberg, L. R., Willett, F. R., & Henderson, J. M. (2024). *A real-time, high-performance brain-computer interface for finger decoding and quadcopter control* (p. 2024.02.06.578107). bioRxiv. https://doi.org/10.1101/2024.02.06.578107

Winkler, I., Haufe, S., & Tangermann, M. (2011). Automatic Classification of Artifactual ICA-Components for Artifact Removal in EEG Signals. *Behavioral and Brain Functions*, *7*(1), 30. https://doi.org/10.1186/1744-9081-7-30

Witt, K., Margraf, N., Bieber, C., Born, J., & Deuschl, G. (2010). Sleep consolidates the effector-independent representation of a motor skill. *Neuroscience*, *171*(1), 227–234. https://doi.org/10.1016/j.neuroscience.2010.07.062

Wodlinger, B., Downey, J., Tyler-Kabara, E., Schwartz, A., Boninger, M., & Collinger, J. (2014). Ten-dimensional anthropomorphic arm control in a human brain-machine interface: Difficulties, solutions, and limitations. *Journal of Neural Engineering*, *12*(1), 016011.

Wolf, P. D., & Reichert, W. M. (2008). Thermal considerations for the design of an implanted cortical brain–machine interface (BMI). *Indwelling Neural Implants: Strategies for Contending with the In Vivo Environment*, 33–38.

Wu, W., Black, M., Gao, Y., Serruya, M., Shaikhouni, A., Donoghue, J., & Bienenstock, E. (2002). Neural decoding of cursor motion using a Kalman filter. *Advances in Neural Information Processing Systems*, *15*.

Wu, W., Shaikhouni, A., Donoghue, J. R., & Black, M. J. (2004). Closed-loop neural control of cursor motion using a Kalman filter. *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, *2*, 4126–4129. https://doi.org/10.1109/IEMBS.2004.1404151

Ye, J., Collinger, J., Wehbe, L., & Gaunt, R. (2023). Neural Data Transformer 2: Multi-context Pretraining for Neural Spiking Activity. *Advances in Neural Information Processing Systems*, *36*, 80352–80374.

Zbinden, J., Sassu, P., Mastinu, E., Earley, E. J., Munoz-Novoa, M., Brånemark, R., & Ortiz-Catalan, M. (2023). Improved control of a prosthetic limb by surgically creating electro-neuromuscular constructs with implanted electrodes. *Science Translational Medicine*, *15*(704), eabq3665. https://doi.org/10.1126/scitranslmed.abq3665

Zhou, K., Yang, Y., Hospedales, T., & Xiang, T. (2020). Deep domain-adversarial image generation for domain generalization. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34*(07), 13025–13032.

Zurawski, E., Behm, K., Dunlap, C., Koo, J., Ismail, F., Boulias, C., Reid, S., & Phadke, C. P. (2019). Interrater Reliability of the Modified Ashworth Scale with Standardized Movement Speeds: A Pilot Study. *Physiotherapy Canada*, *71*(4), 348–354. https://doi.org/10.3138/ptc-2018-0086