

Data sample size needed for prediction of movement distributions

Zachary A. Wright, Moria E. Fisher, Felix C. Huang-IEEE Member and James L. Patton-IEEE Member

Abstract — Human movement ability should be described not only by its typical behavior, but also by the wide variation in capabilities. This would mean that subjects that are encouraged to move throughout their workspace but otherwise free to move any way they like might reveal their unique movement tendencies. In this study, we investigate how much information (data) is needed to reliably construct a movement distribution that predicts an individual’s movement tendencies. We analyzed the distributions of position, velocity and acceleration data derived during self-directed motor exploration by stroke survivors (n=10 from a previous study) and healthy individuals (n=5). We examined whether these simple kinematic variables differed in terms of the amount of data required. We found a trend of decreasing time needed for characterization with the order of kinematic variable, for position, velocity, and acceleration, respectively. In addition, we investigated whether data requirements differ between stroke survivors and healthy. Our results suggest that healthy individuals may require more data samples (time for characterization), though the trend was only significant for position data. Our results provide an important step towards using statistical distributions to describe movement tendencies. Our findings could serve as more comprehensive tools to track recovery in or design more focused training intervention in neurorehabilitation applications.

I. INTRODUCTION

Analysis of the statistical distribution of human movement behavior could offer additional insights to complement clinical assessments [1-3] and conventional engineering metrics. Whereas metrics such as range of motion and accuracy in goal-directed reaching provide a gross overview of movement capabilities, a detailed view of movement tendencies is perhaps more revealing of how the motor system exhibits biases in behavior or even burgeoning skills. Recently, Huang and Patton [4] demonstrated that stroke survivors have uniquely identifiable attributes in the movement distributions during self-directed motor exploration. These unique characteristics may relate to specific motor deficits and provide a “fingerprint” [5] of individual movement capabilities. This points towards the opportunity for customized therapy that better fits each individual.

Before analysis of movement distributions can be

J. L. Patton, M. E. Fisher and Z. A. Wright are with the University of Illinois at Chicago, Illinois 60607 +01(312)413-7664 pattonj@uic.edu and the Rehabilitation Institute of Chicago.

F. C. Huang is with the Sensory Motor Performance Program at the Rehabilitation Institute of Chicago at Chicago, Illinois, 60611.

See <http://smpp.northwestern.edu/Robotics> for more information on this research.

leveraged to design effective training regimens, a better understanding is needed about the best methods of characterization. Studies in training customization have shown benefits of assist-as-needed on a per individual basis [6], focusing on simple reaching movements. Identifying the features of movement distribution specific to an individual’s impairments would facilitate better tracking of recovery and more focused therapeutic interventions. In particular, it remains unclear what differences exist in the observed patterns between healthy and motor impaired individuals. One initial important step is to systematically determine the duration of data required to characterize statistical distributions. While it is possible that all patients may take a different amount of data to be robustly profiled, it is equally possible that a fairly consistent amount of data will reliably give a reproducible profile.

While motor deficits due to stroke cause clear changes in the patterns of preferred movement, it is also unknown what kinematic variables offer the most information about individual characteristics. Stroke survivors exhibit a variety of impairments including spasticity, asymmetry in muscle strength [7], jerky movements [8], abnormal muscle coupling [9], each at varying levels of severity. It is plausible then that in some cases analysis of joint displacement variables may better capture range of motion problems, while in other cases velocity may better capture issues due to spasticity or strength. Consequently, as a preliminary step we set out in this study to investigate if simple kinematic variables differed in terms of the amount of data required to comprehensively characterize individual movement distributions.

In this study we investigate whether particular patterns of movement distributions emerge from analyses of data from stroke survivors and healthy individuals. Our approach draws from conventional cross-validation techniques of machine learning [10] to determine the time required to characterize individual movement patterns. We first examine how the statistical distributions of kinematic variables (position, velocity and acceleration) evolve over continuous motor exploration training. We explore the question of how much data might it take to reliably profile an individual, and whether this differs between stroke survivors and healthy. Our findings provide an important step towards the utilization of statistical distributions to describe movement tendencies.

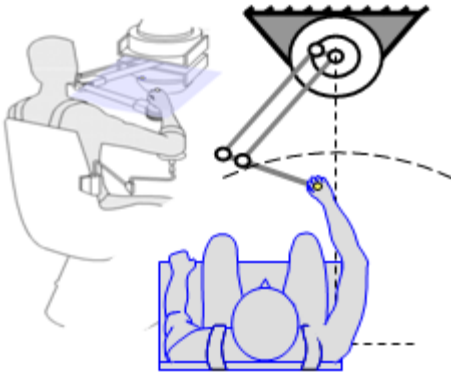


Fig. 1: Subjects freely moved the arm of a robotic device during a motor exploration task.

II. METHODS

A. Experimental Subjects

Ten stroke survivors and five healthy subjects participated in this study. Data for stroke survivors was used from a previous study in which they performed a motor exploration task with and without robot-applied forces accompanied by intermittent trials of a goal-directed circle drawing task. Stroke survivor subjects used their affected arm for all aspects of the experiment. We only consider data during the motor exploration portion in which movement was not interrupted by forces. Healthy subjects completed a similar protocol. Each subject signed a consent form approved by Northwestern University Institutional Review Boards.

B. Experimental Protocol

Subjects controlled the movement of a planar haptics/graphics robotic device (Fig. 1), presented previously [11]. The robot is capable of recording limb position at 200Hz. An overlaying projector provides real-time feedback of the handle position and an animation of two segments approximating the motion of the forearm and upper arm. To focus training on the coordination of the forearm and upper arm, subjects operated the device through a wrist brace.

During the motor exploration task, we instructed subjects to move the robot handle at their own discretion while attempting to acquire various positions, speeds and movement directions within a pre-determined workspace (0.5 x 0.3m). We informed subjects that this task should serve as preparation for a subsequent goal-directed task in which they would perform circular motions. Subjects successfully completed a motor exploration trial once the handle endpoint travelled 25 meters. For each session, subjects completed nine motor exploration trials interspersed with blocks of circle drawing. The time to complete each motor exploration trial varied among subjects since distance travelled depended on endpoint velocities.

C. Analysis

We constructed two-dimensional histograms for three successive state space derivatives (position, velocity and acceleration) in the x and y directions to describe movement

during motor exploration trials. Data was partitioned into 20 equal size bins within specified state space ranges (workspace area for position, -0.7 m/s – 0.7 m/s for velocity, and -7 m/s² – 7 m/s² for acceleration) that were the same across subjects. State space ranges were approximated from the maximum and minimum values reached across the entire subject population. To avoid accumulation of histogram counts within single bins during user-intended periods of rest, we removed data points where the endpoint speed reached below 0.04 m/s and where acceleration reached below 0.07 m/s². To create probability distributions, we then normalized to the total number of observations for a given set of motor exploration data.

To obtain a detailed view of how individual movement patterns changed throughout motor exploration practice, we constructed probability distributions for each kinematic variable of movement data cumulated across time. We divided subject motor exploration datasets into 5s epochs and then partitioned the data within each epoch into two separate datasets: a training dataset and a test dataset which comprised of 75% and 25% of the data, respectively. We constructed separate probability distributions of the training sets for each 5s epoch, combining training data from each successive epoch. We then tabulated a probability distribution of the test data combined from each epoch which was to compare against the training set.

To determine the time required to characterize movement patterns for each state space variable, we calculated the coefficient of determination between the probability distributions of each successive training dataset and the probability distribution of the test dataset. We define *time to characterization* as the time in which the coefficient of determination reached a value of $.95$. We repeated this calculation 50 times using a different random test dataset. We compared the time to movement characterization between each state space variable and between subject populations using a two-factor Analysis of Variance. Differences with a probability of less than $.05$ were considered significant. Bonferroni's correction was used to adjust for multiple comparisons ($.05/9$).

III. RESULTS

For each subject's motor exploration datasets, we constructed 2D probability distributions of position, velocity and acceleration data across cumulative epochs and compared them to the probability distribution of their respective test datasets (Fig. 2). Representative test datasets for each subject were highly predictive of their respective distributions across all motor exploration trials ($R^2 > 0.99$).

Comparisons between the probability distributions of each successive cumulative epoch and respective test datasets revealed a gradual increase in coefficient of determination (Fig. 3). This suggests probability distributions became more similar to representative test datasets with the addition of more data.

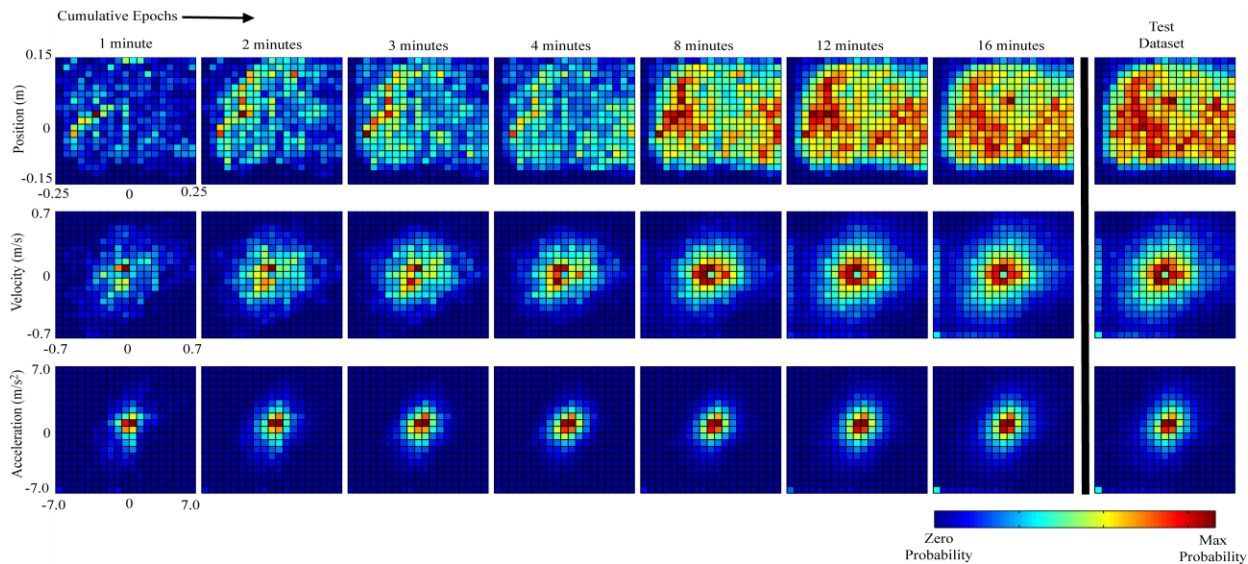


Fig. 2: Representative probability distributions of position, velocity and acceleration data on successive cumulative epochs (5s) compared to the probability distribution of a test dataset (25% of data randomly selected). Probability distributions of motor exploration were constructed from planar movement (fore-aft and left-right axes). Distributions for higher state space derivative resembled respective test datasets at earlier epochs.

For healthy subjects, we detected a 30% difference in time to characterization between position and velocity, a 54% difference between position and acceleration and a 34% difference between velocity and acceleration (Fig. 4A). For stroke survivors, we calculated a 19% difference between position and velocity, a 68% difference between position and acceleration and a 60% difference between velocity and acceleration. A two-factor Analysis of Variance yielded a main effect on selection of state space variable; $F(2, 39) = 12.48, p < .05$. Post hoc comparisons were made using Bonferroni's adjusted alpha level of $.05/9$ (see significance bars on Fig. 4A). These results indicate that the probability distributions of higher state space derivatives (i.e. acceleration) require less time to characterize than lower state space derivatives (i.e. position).

We also found differences in time to movement characterization between healthy and stroke subjects (Fig. 4A). We detected a 44% difference in time to characterization between healthy subjects and stroke

survivors for position, 36% for velocity and 61% for acceleration. A two-factor Analysis of Variance yielded a main effect on subject population; $F(1,39) = 19.06, p < .05$. These results indicate that characterization of state space distributions for stroke survivors can occur earlier than healthy.

In general, stroke subjects completed the motor exploration task (mean \pm SE = 13.5 ± 1.28 minutes) in less time than healthy subjects (mean \pm SE = 20.1 ± 2.30 minutes). Therefore, we tested whether time to characterization was correlated to time to completion (Fig. 4B). Simple linear regression showed a significant correlation for position ($R = .62, p < .05$) and velocity ($R = .36, p = p < .05$), but not for acceleration ($R = .14, p = .16$). This result indicates that the differences in time to characterization between healthy and stroke data may be confounded by the differences in the amount of data available.

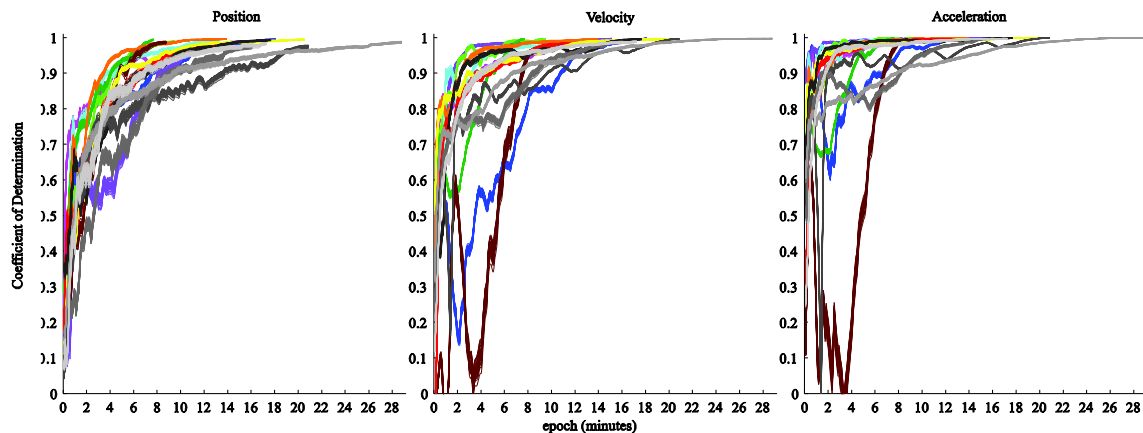


Fig. 3: Coefficient of determination for the probability distribution training as the predictor of the test sets (i.e. 25% of data selected at random) for successive cumulative epochs (1 min intervals) for each subject. Each color represents one subject (healthy - grayscale, stroke - RGB), and each line represents a single computation with a new random training set (50 repeats). These results suggest that more data samples are needed to describe movement distributions for healthy individuals, and position data generally requires more data sample than velocity and acceleration.

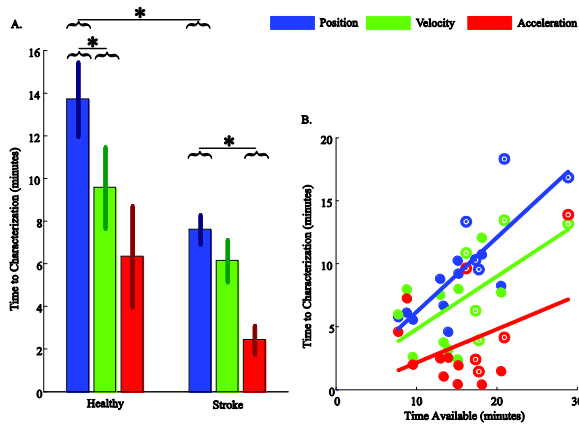


Fig. 4: (A) Duration needed to characterize differences between (1) state spaces and (2) subject populations. Error bars represent standard error. Asterisks represent statistical differences. More data samples are needed to describe lower state space derivatives and healthy subject data. (B) Correlation between completion time and time to characterization. Curves represent best fit line. Closed and open circles represent stroke survivors and healthy subjects, respectively.

IV. DISCUSSION

The purpose of this study was to determine the time required to capture a description of individual movement patterns while subjects performed a motor exploration task. This investigation used a simple statistical approach to determine the time to characterization by comparing probability distributions as they manifested over time. Our first main finding illustrated that higher order kinematic variables require *less* time to characterize. This may point to more effective approaches that consider these higher derivatives. If one considers the fact that we operate in a second order world, creating forces (and torques) from muscle control, it may be no surprise that a more consistent picture of tendencies may come from the derivatives that are proportional to torque via newton's second law.

Our second finding was that healthy subjects' movement patterns surprisingly required *more* time to characterize compared to stroke survivors. While this result may be confounded by the differences in time to completion (hence data available) between the two groups, it is likely informative that the systematic problems with stroke subjects motions are more than random, leading to less time necessary to statistically profile these patients. The fact that we did not observe significant differences across all state space comparisons does not imply that absolute differences in their movement distributions do not exist. Future analysis on movement distributions could help identify specific features unique to each group, which would complement the tools developed in this study. We believe time to characterization could provide a reference for tracking improvement of stroke subjects as they train to expand their movement capabilities.

Time to characterization does not help identify sources of variation in movement distributions, which is evident from the fluctuations that occur in the coefficient of determination (see Fig 3.). We believe there are several factors that contribute to these variations. For instance, at the beginning

of motor exploration subjects are becoming familiar with operating the robot. In contrast, following extensive training, subjects may experience fatigue and a decrease in motivation. Anecdotally, our observations also showed that some subjects exhibit repeated goal-directed movements, for example, repeated circular patterns. Despite these limitations, we observed similar values for time to characterization across subjects. Future analyses may include means to account for these factors that contribute to the variation.

Not only do the tools developed here have implications on neurorehabilitation, they could also apply more generally to a range of human motor behaviors. In particular, this tool might be used to characterize features of motor impairments beyond stroke where clinical assessments are lacking. Time to characterization is likely a general property of statistical distributions that could be used for a variety of domains relating to human behavior; including, electromyography, joint-space variables and electrocorticography.

ACKNOWLEDGEMENT

This work was supported by the National Institute of Health (NIH) under grant R01NS053606 – 05A1.

REFERENCES

- [1] M. Warren, *THE biVABA (Brain Injury Visual Assessment Battery for Adults)*. <http://www.visabilities.com/>; visABILITIES Rehab Services, Inc., 2005.
- [2] C. Gowland, P. Stratford, M. Ward, J. Moreland, W. Torresin, S. Van Hullenaar, *et al.*, "Measuring physical impairment and disability with the Chedoke-McMaster Stroke Assessment," *Stroke*, vol. 24, pp. 58-63, 1993.
- [3] A. R. Fugl-Meyer, L. Jaasko, I. Leyman, S. Olsson, and S. Stegling, "The post-stroke hemiplegic patient. I. a method for evaluation of physical performance," *Scandinavian Journal of Rehabilitation Medicine*, vol. 7, pp. 13-31, 1975.
- [4] F. C. Huang, F. A. Mussa-Ivaldi, C. M. Pugh, and J. L. Patton, "Learning Kinematic Constraints in Laparoscopic Surgery," *IEEE Trans Haptics*, vol. 5, pp. 356-364, Oct 2012.
- [5] Y. Chen, Y. Liu, X. Zhu, H. Chen, F. He, and Y. Pang, "Novel approaches to improve iris recognition system performance based on local quality evaluation and feature fusion," *ScientificWorldJournal*, vol. 2014, p. 670934, 2014.
- [6] J. L. Emken, R. Benitez, and D. J. Reinkensmeyer, "Human-robot cooperative movement training: learning a novel sensory motor transformation during walking with robotic assistance-as-needed," *J Neuroeng Rehabil*, vol. 4, p. 8, 2007.
- [7] C. Mercier, A. M. Bertrand, and D. Bourbonnais, "Differences in the magnitude and direction of forces during a submaximal matching task in hemiparetic subjects," *Exp Brain Res*, vol. 157, pp. 32-42, Jul 2004.
- [8] J. P. Dewald, P. S. Pope, J. D. Given, T. S. Buchanan, and W. Z. Rymer, "Abnormal muscle coactivation patterns during isometric torque generation at the elbow and shoulder in hemiparetic subjects," *Brain*, vol. 118 (Pt 2), pp. 495-510, Apr 1995.
- [9] J. P. Dewald and R. F. Beer, "Abnormal joint torque patterns in the paretic upper limb of subjects with hemiparesis," *Muscle Nerve*, vol. 24, pp. 273-83, Feb 2001.
- [10] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*. Cambridge, MA: MIT Press, 2012.
- [11] F. Gandolfo, F. A. Mussa-Ivaldi, and E. Bizzi, "Motor learning by field approximation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, pp. 3843-6, 1996.