An Extended Framework for Measuring the Information Capacity of the Human Motor System

Teemu Roos and Antti Oulasvirta
Helsinki Institute for Information Technology HIIT
University of Helsinki and Aalto University, Helsinki, Finland
Email: {firstname.lastname}@hiit.fi

***Abstract*—Fitts' law is a fundamental tool in measuring the capacity of the human motor system. It measures information throughput in terms of the tradeoff between the speed and accuracy of aimed movements. However, it is, by definition, confined to prescribed stimulus-response conditions and it leaves out complex skilled performance produced irrespective of the environment. We revisit the information-theoretic basis of Fitts' law with the goal of generalizing it into unconstrained movement. The proposed new metric is based on a subjects ability to accurately reproduce a movement pattern. It can accommodate recorded movement of any duration and composition, and involving contributions of any part(s) of the body. We demonstrate the metric by analyzing publicly available motion capture data. Possible applications include human-computer interaction, sports science, and clinical diagnosis.**

***Index Terms*—Fitts' law, information capacity, stochastic complexity, human motor system, human-computer interaction**

I. Introduction

The purpose of the human motor system is to transform electro-chemical signals in the nervous system into physical movement. The dominant paradigm for studying the information capacity of the human motor system is based on the pioneering work by Paul Fitts in the 1950s [6], [7], [16]. Its primary application is the analysis of user interfaces in human-computer interaction (HCI) [10], [15], [17]; it was, for instance, one of the main drivers in the development and adoption of the computer mouse [3].

Fitts was interested in *aimed movements*; i.e., movement where a pointer (finger, eye fixation, arm, mouse cursor etc.) is moved on top of a spatially expanded target. A good example is moving mouse cursor on top of a button on a computer display. Fitts' law describes the observation that the relationship between movement time MT and spatial characteristics of the required movement is usually very well characterized as:

$$MT = a + b \log_2 \left(1 + \frac{D}{W}\right), \tag{1}$$

where D is the distance from the starting point to the center of the target and W is the width of the target; a and b are empirical parameters that depend on features of the pointing device and the task. MT is typically measured in an empirical procedure involving rapid responses to spatial targets with experimenter-controlled characteristics.

The information-theoretic basis of Fitts' law is centered around the tradeoff between the speed and accuracy of movements produced by the motor system. In information-theoretic terms, the *capacity* of the motor system as a channel of communication is limited by this tradeoff. It is important to note that measuring only the speed of movements is not sufficient because, as Fitts points out,

> *" [s]ince measurable aspects of motor responses, such as their force, direction, and amplitude, are continuous variables, their information capacity is limited only by the amount of statistical variability, or noise, that is characteristic of repeated efforts to produce the same response. " [6]*

The information theoretic interpretation of Fitts' law [6], [7], [9], [16], [17], where the information throughtput is defined by the channel capacity, has been immensely popular since it enables the comparison of performance across situations with different characteristics. The *index of performance* (IP) defines the information throughput in units of *bits per second* (bps):

$$IP = 1/b. \tag{2}$$

IP is argued to be a good metric because it stays relatively constant over a broad range of target characteristics [15], [17], providing a natural basis for comparison of pointing devices. The mouse, for example, typically reaches ca. 4 bps, and joystick ca. 2 bps [15].

The motivation for the present work is that we believe that important aspects of the information potential of human motor system are not covered by the Fitts' law paradigm, and that consequently, the capacity of human motor system is systematically underestimated. Fitts' law, by definition, is constrained to aimed movements in target conditions *prescribed by the environment*. This leads to three shortcomings. Firstly, the "information" that is measured is tantamount to the subject's ability to motorically conform to extrinsic constraints, excluding entirely action, i.e., action produced irrespective of its absolute position in respect to perceivable environmental constraints. Such movements are important in many skilled activities, such as conducting and dancing. The issue of underestimation is exacerbated by the empirical

This work has been submitted to the 2011 IEEE Information Theory Workshop (ITW 2011), Paraty, Brazil.

paradigm, which utilizes very simple repetitive movements with simple trajectories. Seconly, Fitts' law does not account for information in *simultaneous movement of multiple body parts* (for an exception, see [12]). There are 640 muscles, 200-300 joints, and 206 bones in the human body. Obviously we are not able to independently control each of them, but some separation is possible; for instance, the thumb and the index finger can be moved relatively independently of each other and the three other fingers [8]. Thirdly, most skilled activities involve compound tasks, with multiple aimed and other types of movement performed simultaneously and sequentially. Due to these three limitations, we argue that the Fitts' law paradigm is not suitable for the study of skilled motor action in unconstrained domains; i.e., precisely the ones that can be expected to contain the most information!

Extending Fitts' definition, we define information capacity in terms of the ability to accurately *reproduce any previously performed movement pattern*. An infant is a good example. At any moment in time, the infant's movement can appear complex, but the fact that he or she cannot reproduce it at will means that the motor system lacks information capacity.

Our formulation is based on subjects performing arbitrarily complex un-prescribed movements; Fitts' paradigm, involving only experimenter-defined pointing tasks, is a special case. The formulation can accommodate movement of any duration and composition and involving contributions of any part of the body.

The rest of the paper is organized as follows. In Sec. II, we describe a measure of shared information between two movement sequences.[1] The data and the preprocessing steps are detailed in Sec. III, and the results of the experiments are summarized in Sec. IV. To conclude, in Sec. V we discuss potential applications and outline future work.

II. Information Measure

To quantify the information capacity, it is necessary to separate the *controlled* aspects of the performed sequence of movements from the *unintentional* aspects that are unavoidably present in all motor responses. As discussed above, the strictly defined range of admissible performances in Fitts' paradigm has a similar function: it rules out apparently complex, uncontrolled (random) sequences of movements. Instead of restricting the allowed movements, we propose to solve this task by having a sequence *repeated* as exactly as possible by the same subject. This makes it possible to obtain an estimate of the variability of the two patterns, and subtract the complexity (entropy) due to it from the total complexity of the repeated performance. In other words, information is measured by two aspects of the performance: i) the complexity of a movement pattern, and ii) the precision with which it can be repeated.

For simplicity, we start by treating the one-dimensional case where movements are characterized by a single measurement per time frame. Let $\mathbf{x} = x_{-1}, \ldots, x_n$ denote a sequence where x_t gives the value of the measured feature at time $t \in \{-1, \ldots, n\}$. We start the sequence from x_{-1} instead of x_1 for notational convenience: the first two entries guarantee that an autoregressive model with a look-back (lag) of two steps can be fitted to exactly n data points.

In order to define the complexity of $\mathbf{x}$, we fit a second-order autoregressive model

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \epsilon_t, \tag{3}$$

where $\beta_0, \beta_1, \beta_2$ are real-valued parameters to be tuned using least squares, and ϵ_t are assumed to be zero mean i.i.d. Gaussian errors. The second-order model accounts for the basic physical principle that once the movement vector (including direction and velocity) is specified, constant movement contains no information whatsoever. The complexity of $\mathbf{x}$ is determined by the residuals

$$r_t = x_t - \hat{x}_t = x_t - (\hat{\beta}_0 + \hat{\beta}_1 x_{t-1} + \hat{\beta}_2 x_{t-2}),$$

where $\hat{x}_t$ denotes the predicted value based on the least squares estimates $\hat{\beta}_0$,$\hat{\beta}_1$,$\hat{\beta}_2$.

To quantify the complexity, we apply the classic two-part approximation of Rissanen's *stochastic complexity* [13]:

$$SC(\mathbf{x}) = \frac{n}{2} \log_2(2\pi e \hat{\sigma}^2) + \frac{k}{2} \log_2 n, \tag{4}$$

where $\hat{\sigma}^2 = \sum_{t=1}^{n} r_t^2 / n$ is the residual variance and $k = 4$ denotes the number of parameters, including the variance; $\log_2$ denotes the binary logarithm. The first term, which is equal to n times the differential entropy of a Gaussian density $\mathcal{N}(0, \hat{\sigma}^2)$, can be interpreted as the ideal code-length of the residual sequence under the fitted model, see [4]. The second term, which increases with model complexity (measured by the number of parameters), will act to balance the complexity of the model and its ability to fit the data, see also [5].

In order to obtain a meaningful measure of the extent to which the sequence $\mathbf{x}$ is controlled, we consider a second sequence $\mathbf{y} = y_1, \ldots, y_n$, of the same length as $\mathbf{x}$. The idea is that if sequence $\mathbf{y}$ is similar enough to $\mathbf{x}$, we can improve the fit of the simple autoregressive model, Eq. (3), by including the t'th observation of $\mathbf{y}$ as a regressor in addition to x_{t-1} and x_{t-2}:

$$x_t = \eta_0 + \eta_1 x_{t-1} + \eta_2 x_{t-2} + \eta_3 y_t + \varepsilon_t, \tag{5}$$

where $\eta_0, \ldots, \eta_3$ are real-value parameters, and ε_t are again zero mean i.i.d. Gaussian errors.

We denote the residuals remaining after a least squares fit by

$$s_t = x_t - \hat{x}_t(y_t) = x_t - (\hat{\eta}_0 + \hat{\eta}_1 x_{t-1} + \hat{\eta}_2 x_{t-2} + \hat{\eta}_3 y_t),$$

where $\hat{x}_t(y_t)$ is the predicted value taking into account the side-information y_t. The stochastic complexity of $\mathbf{x}$ conditioned on $\mathbf{y}$ is given by

$$SC(\mathbf{x} \mid \mathbf{y}) = \frac{n}{2} \log_2(2\pi e \hat{\sigma}^2(\mathbf{y})) + \frac{k'}{2} \log_2 n, \tag{6}$$

[1] The R code required to implement our method is available at http://www.cs.helsinki.fi/teemu.roos/R/infocapacity.R.

where $\hat{\sigma}^2(\mathbf{y})$ denotes the residual variance of the residuals s_t; we now have $k' = 5$ due to the additional parameter in the model compared to Eq. (3).

The difference between the stochastic complexity $SC(\mathbf{x})$, and the conditional complexity $SC(\mathbf{x} \mid \mathbf{y})$ gives the *shared information*[2] of $\mathbf{x}$ and $\mathbf{y}$:

$$\begin{aligned} SI(\mathbf{x}\,;\,\mathbf{y}) &= SC(\mathbf{x}) - SC(\mathbf{x} \mid \mathbf{y}) \\ &= \frac{n}{2}\log_2\left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2(\mathbf{y})}\right) + \frac{1}{2}\log n, \end{aligned} \tag{7}$$

where the last terms is due to the additional parameter, $k' = k + 1$. Unlike the differential entropy, the shared information has a direct interpretation in terms of the reduction in *bits* required to encode the sequence $\mathbf{x}$ due to the side information $\mathbf{y}$ being available. Since the shared information in $\mathbf{x}$ and $\mathbf{y}$ excludes, with high probability, most of the uncontrolled movements and inaccuracies, we argue that it provides a measure of the controlled information in $\mathbf{x}$. To achieve high shared information, a movement has to be both complex and accurately controlled so that it can be repeated with precision.

Finally, we define the *throughput* in a sequence $\mathbf{x}$ conditioned on sequence $\mathbf{y}$ as the shared information per second:

$$TP(\mathbf{x} \mid \mathbf{y}) = \frac{R\,SI(\mathbf{x}\,;\,\mathbf{y})}{n}, \tag{8}$$

where R denotes the frame rate (frames per second).

To handle k-dimensional sequences, where each time frame x_t is composed of k measured components (features), $x_t = (x_t^{(1)}, \ldots, x_t^{(k)})$, we simply sum up the shared information of each of the components separately. We note that this is likely to exaggerate the throughput as redundant information that is contained in more than one component is counted several times. Hence, the present formulation can only be expected to yield results that are *qualitatively* consistent (see Sec. IV). In future work, we plan to explore ways to correct this bias.

III. Data and Preprocessing

In order to study unconstrained performances without limiting ourselves to specific tasks or parts of the body, we consider motion capture data. Motion capture data is typically obtained by recording a subject by a set of cameras, and using special-purpose image processing technologies to convert the recorded video into variables such as locations and angles of joints (wrists, elbows, shoulders, waist, knees, etc).

For a proof-of-concept study, we downloaded all the 20 motion capture sequences in the category *Subject #5: modern dance* from the CMU Motion Capture Library[3], see Table I and Fig. 1. Some of the sequences are seemingly more complex than others, and some pairs of sequences are similar enough to be considered repetitions of one another. Intuitively, it is expected that the most accurate repetitions of complex patterns reach the highest throughput values. The sequences are recorded at frame rate 120 per second. For each frame, the data contains the measured angles of 62 different angle features, derived from 40 infrared cameras. We removed features with zero or nearly zero variance (features 1,3,25,26,34,37,38,46) in order to avoid indeterminacies in the fitting process.

TABLE I

Summary of the data used in the experiments. Sequences referred to in Sec. IV are emphasized.

#	Label	n
1	*walk*	598
2	*expressive arms, pirouette*	1123
3	**sideways arabesque, turn step, folding arms**	434
4	**sideways arabesque, folding arms, bending back**	1199
5	*quasi-cou-de-pied, raised leg above hip-height, jete en tourant*	915
6	*cartwheel-like start, pirouettes, jete*	885
7	*small jetes, attitude/arabesque, shifted-axis pirouette, turn*	1191
8	*rond de jambe in the air, jete, turn*	721
9	*glissade devant, glissade derriere, attitude/arabesque*	1143
10	*glissade devant, glissade derriere, attitude/arabesque*	817
11	*sideways steps, pirouette*	591
12	*arms held high, pointe tendue a terre, upper body rotation*	1354
13	*small jetes, pirouette*	1095
14	**retire derriere, attitude/arabesque**	642
15	**retire derriere, attitude/arabesque**	540
16	*coupe dessous, jete en tourant*	525
17	*coupe dessous, grand jete en tourant*	1043
18	**attitude/arabesque, jete en tourant, bending back**	1829
19	**attitude/arabesque, jete en tourant, bending back**	860
20	*attitude/arabesque, jete en tourant, bending back*	1095

The inherent problem in predicting one motion sequence by another is the possible (and very common) misalignment of the sequences in time. Usually, even very carefully repeated movements are slightly out of synchronization, and hence when predicting the t'th frame of sequence $\mathbf{x}$, the most useful frame of sequence $\mathbf{y}$ may not be the t'th frame but the $t+\delta$'th one with $\delta \neq 0$. Therefore, it is necessary to align the two sequences to obtain a better synchronization.

We aligned each pair of sequences in the data set by applying Canonical Time Warping (CTW)[4] [18], a state-of-the-art technique for aligning sequences describing human behavior. CTW uses the more traditional Dynamic Time Warping (DTW) [11] as an initial solution but improves it by adopting features from Canonical Correlation Analysis (CCA) (see [2]). This allows alignment based on a more flexible concept of similarity than usually used in DTW. Since the implementation provided by the authors (of [18]) had problems handling all the features, we based the alignment on the first 25 features which seemed to capture the synchronization properties of the signals sufficiently well.

The result of a pairwise alignment of two sequences, with possibly different lengths, is a new pair of aligned sequences whose lengths are equal, such that each frame in one sequence matches as well as possible with the same movement (similar

[2]Clearly, the shared information is closely related to *mutual information*, $I(\mathbf{x}\;;\;\mathbf{y}) = h(\mathbf{x}) - h(\mathbf{x} \mid \mathbf{y})$, where $h(\cdot)$ denotes the (differential) entropy. However, even if the two concepts are fundamentally similar, it is important to note that the shared information is not necessarily symmetric in the two arguments, can be negative, and lacks other key properties of mutual information.

[3]Available at http://mocap.cs.cmu.edu. The Library was supported by NSF EIA-0196217.

[4]Matlab code is available at www.humansensing.cs.cmu.edu/projects/ctwCode.html.

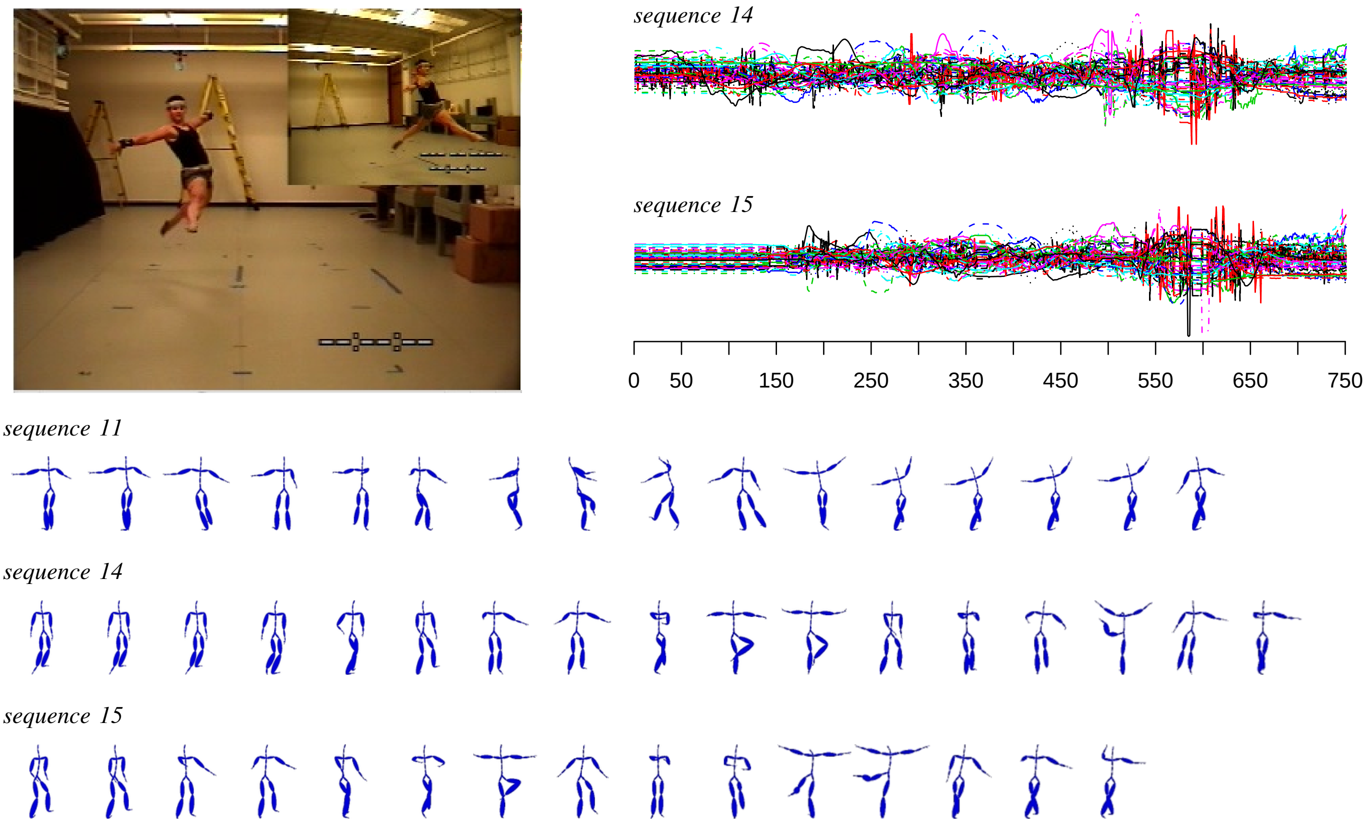

Fig. 1. Data from the CMU Motion Capture Library. TOP-LEFT: An example of a motion capture situation on video. TOP-RIGHT: The plotted sequences of two motion capture sequences after alignment — note the constant stretch in the first 150 frames of the lower sequence caused by the alignment. BOTTOM: Three examples of animated motion capture sequences (unaligned). Each figure shows a frame at rate three frames per second, see Table I.

measured features) in the other. To achieve this, the CTW algorithm duplicates some of the frames in each sequence so as to "slow down" the sequence in question at suitable points; see the example in Fig. 1. When measuring the throughput, we skip the duplicated frames in order to avoid unnecessarily magnifying their impact. Hence, if frame t is duplicated in sequence $\mathbf{x}$ so that in the aligned sequence, $\mathbf{x}'$, frames t and $t+1$ are identical, we skip the $t+1$'th frame (of both $\mathbf{x}'$ and $\mathbf{y}'$) when evaluating the throughput, Eq. (8). The sequences were also normalized so that each feature has mean zero and unit variance.

An unwanted consequence of the use of alignment methods in preprocessing the motion capture data is the loss of timing information. Clearly, a significant amount of controlled information are required for timing the motor responses. Working with aligned sequences, there is no way to measure the accuracy to which the repeated performance is synchronized with the original performance. One possibility is to examine the alignment itself to see how much information is required to bring the two sequences in close agreement, and to add this information to the information content due to spatial accuracy.

IV. Results and Discussion

Of all the pairwise throughput values, $C(\mathbf{x} \mid \mathbf{y})$, the highest one, 103 bits per second (bps), is obtained for sequence 15 conditioned on sequence 14, see Fig. 1. Their similarity is easily confirmed visually from the video recordings and the animated reconstructions available in the CMU Motion Capture Library. The second and third highest values are achieved respectively by sequence 3 conditioned on sequence 4 (36 bps), and sequence 18 conditioned on sequence 19 (29 bps). The fourth highest value, 14 bps, is obtained by the same pair as the highest value, in the other order, sequence 14 conditioned on sequence 15. In all cases, the similarity is can be visually confirmed from the video or animation sequences. The remaining values are all clearly below 10 bps, and in fact, most of them are negative[5]. This is reasonable since most of the sequences cannot be considered as being repetitions of the others at all.

The maximum throughput value, 103 bps, is very likely to be overestimated due to reasons discussed in Sec. II. However, the relative magnitude of the values appears to consistently reflect the similarity of the sequences. One thing that stands out in the results is the tendency of shorter sequences to

[5]Negative values are possible due to the second term, $\frac{1}{2}\log_2 n$, in Eq. (7). In terms of the Minimum Description Length (MDL) Principle [5], [13], this would be taken to indicate that the model *without* y_t as one of the independent variables, Eq. (3) is superior to the model that includes y_t, Eq. (5). This is equivalent to model selection using the Bayesian Information Criterion (BIC) [14].

achieve higher throughput. This requires further study but it is likely explained, at least to some extent, by the higher information content *per second* in more rapid performances. In this light, it is even more curious that for the pair 18–19 the longer sequence (18) achieves a significantly higher throughput.

The data we used in our experiment serves to provide a proof-of-concept, even though the data we used was not specifically designed for our purpose. The pairs of sequences for which the throughput values turn out to be non-negligible (and non-negative) are clearly very similar. In the future, we will collect data by having subjects repeat performances as accurately as possible, thus maximizing the shared information between the movement sequences. Subjects whose information capacity is expected to be exceptional are of specific interest: on one hand, artists and athletes, and on the other hand, children, elderly people, and people with disabilities.

V. Conclusions and Future Work

The experiment we have described demonstrates the main idea in our framework, i.e., extending the prevailing information-theoretic framework to allow completely unconstrained movements, and thereby, to determine the maximum of the achievable information capacity. Motion capture data provides the best way to characterize such movements in a way that does not rule out any potentially informative aspects in them.

That said, it will be interesting to compare the capacity estimates obtained by other methods, such as pointing devices (the traditional tool in Fitts' paradigm), data gloves, etc., and to see if the earlier results are replicated. For instance, it is interesting to see if more information can be extracted from Fitts' original reciprocal pointing task by recording the movements by a data glove or motion capture: the question is whether the path along which the hand operating the pointer moves between the two targets carries additional information beyond the information provided by the end-points, and if yes, how much.

The formalism we presented is one specific way to measure the amount of controlled information in movement sequences. Other, probably more accurate measures can be derived by applying state-of-the-art universal modeling techniques, see e.g. [5] and references therein. More specifically, it is necessary to handle the correlations between the different features, as pointed out in Sec. II. This will also tell us how much *independent* (non-redundant) information can be simultaneously expressed by different parts of the motor system. Likewise, a method to take into account higher level regularities, such as repetitions, in the movement sequences. Compression algorithms such as Lempel-Ziv encoding may be helpful in this task.

Finally, achieving the goal of constructing a complete and reliable measure of information capacity will lead to a wealth of useful knowledge about the human motor system. We are currently carrying out preliminary experiments with several categories in the CMU Motion Capture Library, such as those related to various sports (football, basketball, American football). It will be exciting to be able to, even in principle, answer questions like: who expresses most controlled information, a ballet dancer performing *Swan Lake*, an expert pianist, or Lionel Messi dribbling through the defence lines[6]. More concrete utility is to be seen, for instance, in the study of novel human-computer interfaces that involve free whole-body expression. Possible applications in sports science include training of complex motor schemas with reference models. Potential new diagnostic tools based on monitoring changes in the information capacity of the motor system may offer great societal value through early identification of neurological disorders related to motor dysfunction and in monitoring recovery of neuroplasticity after lesions.

References

[1] J. Accott and S. Zhai. "Performance evaluation of input devices in trajectory-based tasks: an application of the steering law", *Proc. CHI'07*, ACM Press, pp. 466–472, 1999.

[2] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*, Wiley, 2003.

[3] P. Atkinson. "The best laid plans of mice and men: the computer mouse in the history of computing", *Design Issues* **23**:46–61, 2007.

[4] T. Cover and J. Thomas. *Elements of Information Theory*, 2nd Ed., Wiley, 2006.

[5] P. Grünwald. *The Minimum Description Length Principle*, MIT Press, 2007.

[6] P. M. Fitts. "The information capacity of the human motor system in controlling the amplitude of movement", *J Experim Psychology* **47**:381–391, 1954.

[7] P. M. Fitts and J. R. Peterson. "Information capacity of discrete motor responses", *J Experim Psychology* **67**:103-112, 1964.

[8] L. A. Jones and S. J. Lederman. *Human Hand Functioning*, Oxford University Press, 2006.

[9] I. S. MacKenzie. "A note on the information-theoretic basis for Fitts' law", *J Motor Behavior* **21**:323-330, 1989.

[10] I. S. MacKenzie. "Fitts' law as a research and design tool in human-computer interaction", *Human-Computer Interaction* **7**:91-139, 1992.

[11] L. Rabiner and B.-H. Juang. *Fundamentals of Speech Recognition*, Prentice Hall, 1993.

[12] G. H. Robinson and R. C. Kavinsky. "On Fitts' law with two-handed movement", *IEEE Trans Syst, Man & Cybern*, **6**:504–505, 1976.

[13] J. Rissanen. "Modeling by shortest data description", *Automatica* **14**:445–471, 1978.

[14] G. Schwarz. "Estimating the dimension of a model," *Annals of Statistics* **6**:461–464, 1978.

[15] R. W. Soukoreff and I. S. MacKenzie. "Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI", *Int J Human-Computer Studies* **61**:751–789, 2004.

[16] A. T. Welford. *Fundamentals of Skill*, Methuen, 1968.

[17] S. Zhai. "On the validity of throughput as a characteristic of computer input," *IBM Research Report RJ 10253*, IBM Research Center, Almaden, California, 2002.

[18] F. Zhou and F. de la Torre. "Canonical time warping for alignment of human behavior", *Advances in Neural Information Processing Systems* (NIPS), 2009.

[6]Especially the latter performance, of course, is never exactly repeated as a whole. The performance is, nevertheless, composed of smaller units that are likely to stay the same or very similar over different repetitions.