Modeling the speed-accuracy tradeoff using the tools of information theory

Julien Gori

université PARIS-SACLAY

NNT : 2018SACLT022

Thèse de doctorat

Modeling the Speed-Accuracy Tradeoff Using the Tools of Information Theory

Thèse de doctorat de l'Université Paris-Saclay
préparée à Télécom ParisTech

Ecole doctorale n°580 Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat: Informatique

Thèse présentée et soutenue à Paris, le 20 Décembre 2018, par

JULIEN GORI

Composition du Jury :

Caroline Appert Directeur de Recherche, CNRS, LRI	Présidente
Gilles Bailly Chargé de Recherche, CNRS, ISIR	Rapporteur
Antti Oulasvirta Associate Professor, Aalto University	Rapporteur
Pablo Piantanida Associate Professor, CentraleSupélec	Examinateur
Stéphane Huot Directeur de Recherche, Inria Lille	Examinateur
Jörg Müller Full Professor, Bayreuth University	Examinateur
Olivier Rioul Professeur des Universités, Télécom ParisTech	Directeur de thèse
Yves Guiard Professeur Emérite, Télécom ParisTech	Co-directeur de thèse
Michel Beaudouin-Lafon Professeur des Universités, Université Paris-Sud	Invité

Acknowledgments

I would firstly like to thank all the researchers that have shared their experimental data with me, helping me save valuable time: I am indebted to Olivier Chapuis, Michel Beaudouin-Lafon, Renaud Blanch, Michel Ortega, Yves Guiard, Simon Perrault, Quentin Roy, Halla Olafsdottir, Bruno Fruchard. I also wish to thank all of those who published their empirical data openly, and which I have used at some point in my thesis: Ken Goldberg, Siamak Faridani, Ron Alterowitz, Alvin Jude, Darren Guiness, Michael Poor, Jörg Müller, Antti Oulasvirta, Roderick Murray-Smith, Krzysztof Gajos, Katharina Reinecke, Charles Hermann. I can't stress enough the importance of releasing empirical data. I also wish to thank Alexandra Elbakyan.

Secondly, I would like to thank my two supervisors Olivier Rioul and Yves Guiard. Their collaboration started around 2011 when Yves, the experimental psychologist of the computer science department knocked on the door of Olivier, the applied mathematician from the digital communications team, to discuss information theory and Fitts' law. Eventually, they would offer a PhD position, which I gladly took in 2015. Throughout the three years and three months that it lasted, I have taken advantage many times of Yves' experience on the topic. I think his willingness and audacity to challenge assumptions and presumptions, even trivial ones, has somewhat rubbed off on me. I am heavily indebted to Olivier's ideas on the feedback scheme, as well as the insights he has given on many occasions: Olivier has the ability to turn complicated matters into simple ones. Both Olivier and Yves have given me the freedom to pursue more or less whatever I wanted, while still providing guidance when needed, and for this I am grateful. I also wish to acknowledge Michel Beaudouin-Lafon, with whom I have collaborated multiple times.

Je salue aussi les collègues avec qui j'ai eu l'occasion de faire de l'enseignement, aussi bien en informatique (Etienne et l'équipe d' INF104) qu'en électronique (Patricia et l'équipe d'ELEC 101, et en particulier Chaddi avec qui j'ai concocté et encadré plusieurs projets d'élèves) et tout le personnel de Telecom, en particulier Yvonne et Chantal. Je rends aussi un hommage à tous mes amis, de Telecom (ComElec et DIVA) ou d'ailleurs (anciens du M2-SAR, anciens de Cachan, EEA et footeux) sans qui l'expérience de la thèse aurait été bien moins plaisante.

Je réserve mon dernier remerciement (en français) à Elsa qui m'a toujours soutenue, et qui n'a jamais réchigné à relire mes papiers, même sous leur forme la moins enthousiasmante, ni à servir de cobaye pour mes présentations, souvent décousues.

I would finally like to thank the reviewers Antti Oulasvirta and Gilles Bailly for the time they have taken to read through and comment this thesis.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences et technologies de l'information et de la communication (STIC)

Titre : Modélisation du Compromis Vitesse Précision à l'aide des Outils de la Théorie de l'Information

Mots clés : Compromis vitesse précision, loi de Fitts, théorie de l'information de Shannon

Résumé : La loi de Fitts, qui relie le temps de mouvement MT dans une tache de pointage aux dimensions de la cible visée D et W est usuellement exprimée à partir d'une imitation de la formule de la capacité de Shannon $\mathrm{MT} = a + b \log_2 (1 + \mathrm{D}/\mathrm{W})$. Toutefois, l'analyse actuelle est insatisfaisante: elle provient d'une simple analogie entre la tache de pointage et la transmission d'un signal sur un canal bruité sans qu'il n'y ait de modèle explicite de communication.
Je développe d'abord un modèle de transmission pour le pointage, où l'indice de difficulté $\mathrm{ID} = \log_2 (1 + \mathrm{D}/\mathrm{W})$ s'exprime aussi bien comme une entropie de source et une capacité de canal, permettant ainsi de réconcilier dans un premier temps l'approche de Fitts avec la théorie de l'information de Shannon. Ce modèle est ensuite exploité pour analyser des données de pointage récoltées lors d'expérimentations contrôlées mais aussi en conditions d'utilisations réelles.
Je développe ensuite un second modèle, focalisé autour de la forte variabilité charactéristique du mouvement humain et qui prend en compte la forte diversité des mécanismes de contrôle du mouvement: avec ou sans voie de retour, par intermittence ou de manière continue. À partir d'une chronométrie de la variance positionelle, évaluée à partir d'un ensemble de trajectoires, on remarque que le mouvement peut-être découpé en deux phases: une première où la variance augmente et une grande partie de la distance à couvrir est parcourue, est suivie d'une deuxième au cours de laquelle la variance dimininue pour satisfaire les contraintes de précision requires par la tache. Dans la deuxième phase, le problème du pointage peut-être ramené à un problème de communication à la Shannon, où l'information est transmise d'une "source" (variance à la fin de la première phase) à une "destination" (extrémité du membre) à travers un canal Gaussien avec la présence d'une voie de retour. Je montre que la solution optimale à ce problème de transmission revient à considérer un schéma proposé par Elias. Je montre que la variance peut décroitre au mieux exponentiellement au cours de la deuxième phase, et que c'est ce résultat qui implique directement la loi de Fitts.

Title : Modeling the Speed-Accuracy Tradeoff Using the Tools of Information Theory

Keywords : Speed-accuracy tradeoff, Fitts' law, Shannon's information theory

Abstract : Fitts' law, which relates movement time MT in a pointing task to the target's dimensions D and W is usually expressed by mimicking Shannon's capacity formula $\mathrm{MT} = a + b \log_2 (1 + \mathrm{D}/\mathrm{W})$. Yet, the currently received analysis is incomplete and unsatisfactory: it stems from a vague analogy and there is no explicit communication model for pointing.
I first develop a transmission model for pointing tasks where the index of difficulty $\mathrm{ID} = \log_2 (1 + \mathrm{D}/\mathrm{W})$ is the expression of both a source entropy and a channel capacity, thereby reconciling Shannon's information theory with Fitts' law. This model is then leveraged to analyze pointing data gathered from controlled experiments but also from field studies.
I then develop a second model which builds on the variability of human movements and accounts for the tremendous diversity displayed by movement control: with of without feedback, intermittent or continuous. From a chronometry of the positional variance, evaluated from a set of trajectories, it is observed that movement can be separated into two phases: a first where the variance increases over time and where most of the distance to the target is covered, followed by a second phase where the variance decreases until it satisfies accuracy constraints. During this second phase, the problem of aiming can be reduced to a Shannon-like communication problem where information is transmitted from a "source" (variance at the end of the first phase), to a "destination" (the limb extremity) over a "channel" perturbed by Gaussian noise with a feedback link. I show that the optimal solution to this transmission problem amounts to a scheme first suggested by Elias. I show that the variance can decrease at best exponentially during the second phase, and that this result induces Fitts' law.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France

Introduction

Presently, most interfaces are graphic (so-called Graphical User Interfaces, or GUIs) i.e., interaction between the user and the computing device is based on visual graphics produced by the latter and displayed on a screen the users respond or act upon. Current trends in interaction techniques show an increasing reliance on modalities of vision and touch as shown e.g., by the evolution of mobile phones towards ever larger screens and a switch from keyboard-typing to touch typing.

Aimed movement is an essential component of many Graphical User Interfaces. For example, when using a tablet, commands are selected by touching a specific area, say a rectangle labeled "minimize" should the user wish to hide a window. The movement produced by the user to reach the rectangle is an instance of what experimental psychologists call "(voluntary) aimed movements". These are rather complex movements that require significant practice to guarantee their precision and efficiency. A long learning process is actually required to turn our childhood's imprecise saccades into the continuous and seemingly effortless movements that we are accustomed to. Aimed movements, also referred to as goal-directed movements, pointing movements or reaching movements, are thus much more complex, and should be distinguished from the simple rhythmic movements or the even simpler reflex movements.

Aimed movements have been formally studied for more than a century, and it has long been known that one can deliberately slow down one's movement to achieve a better precision. This so-called speed-accuracy tradeoff has notoriously been quantified by an experiment proposed in 1954 by Paul Fitts [38], where a participant has to aim for a target of size W located at a distance D, and movement time MT is measured from motion start to motion end. As one expects, MT increases with D and decreases with W, but surprisingly, the relevant parameter to predict movement time is the *relative accuracy* $\mathrm{D/W}$. The so-called Fitts' law reads:

$$\mathrm{MT} = a + b\ \log_2\big(1 + \mathrm{D}/W\big), \tag{1}$$

where a and b are empirical parameters to be estimated from experimental data.

In Windows, Icons, Menus, Pointer (WIMP) systems, such as those typically found on mainstream operating systems, the user moves a cursor (pointer) with an input device, usually a mouse. In a seminal study of 1978 on the evaluation of pointing devices in Human Computer Interaction (HCI), Card et al. [11] found that Fitts' law models cursor movements accurately. Fitts' law has since then been used extensively and successfully in HCI. Hence, Fitts' law is not only useful for direct pointing movements, such as those produced on a tablet, but also to virtually any indirect technique that mobilizes a cursor.

There are three scenarios in HCI where evaluating pointing performance with Fitts' law is used:

- When comparing two devices against each other e.g., a mouse versus a joystick, pointing performance can be leveraged to put one technique above the other. In

fact, the study by Card et al. is often credited for being a reason why the mouse was chosen as a standard input device on desktop computers [138].

- The ease with which one can interact with an interface can be evaluated and/or predicted from the pointing performance [142]. For example, the average movement time for selecting an item in a menu can be computed. Thus, Fitts' law is a tool that helps, or even guides the design of interfaces.

- A user's performance can be monitored by measuring his pointing capabilities, providing useful clues about him [76].

Another reason for the importance of Fitts' law in HCI has not so much to do with practical concerns, but more with what Fitts' law tells about HCI itself. Fitts' law is one of the few quantitative formulas in HCI, and probably as close as one gets to a formal law, such as those found in the so-called "hard sciences". For those that advocate for a "hardening" of HCI [50, 112], Fitts' law is certainly essential.

This is particularly relevant, seeing that the theoretical explanation of Fitts' law has been challenged on many occasions, and that the resulting theoretical *flou* has lead to several competing formulations and metrics to characterize the same concepts, such as e.g. *throughput*. Most of the insights gained in HCI are the outcome of empirical studies [165] — this includes Fitts' law studies. This comes with many disadvantages, such as the time and effort required to run a study, as well as all the risks that come with running experiments on participants that are often scarce, easily tired and distracted.

The main goal of this thesis is to propose a theoretical model for Fitts' law that is more satisfying than the current state of the art. The hope is not only to end up with a more formal model for pointing that will better support the law's theoretical basis, but also to provide new practical solutions. An obvious deficiency of Fitts' law, apart from the incongruities described above is that, from the whole trajectory, it considers only the starting and stopping. This can be improved, as a model describing the whole trajectory is presented in Chapter 6.

The mathematical framework used for this work is Shannon's information theory [132, 18], a mathematical tool that mostly serves to compute achievable transmission rates in digital communications. Although using information theory within the context of aimed movements might seem surprising at first, there are in fact several good reasons to adopt the framework.

The first is historical as Fitts' law was originally "derived" by analogy with Shannon's information theory—as a matter of fact $\log_2\left(1+\mathrm{D}/\mathrm{W}\right)$ in (1) is usually expressed in bits. One of the challenges of this thesis is thus to obtain Fitts' law through a formal analysis rather than through Fitts' vague analogy.

The second reason is that information-theoretic concepts come quite naturally, if one is ready to consider the stochastic nature of a *whole set* of pointing trajectories rather than single movements. It turns out that aimed movements and HCI offer an unexpected playground for information-theoretic tools.

A final reason to consider information theory comes from the following observation. Many explanations for Fitts' law (and movement control in general) are derived from dynamical considerations. One then gets the sense that "if only we had better actuators and faster sensors, motor control would be trivial" [147]. However, some of the oldest empirical results already show that what limits movement time is probably not so much linked to the dynamics of the system: Fitts' early experiments [38] show almost no differences when the participants use a light or a heavy stylus, and Fenn [36] commented that "muscles viewed only as sources of energy" are capable of moving the limbs at much faster rates than those typically observed. Hence, considering information rather than energy could prove insightful.

It should by now be apparent that this thesis is multidisciplinary. The information-theoretic framework is familiar mostly to electrical engineers, whereas aimed movements are traditionally studied by other fields, such as experimental and behavioral psychology, neurology, and ergonomics and human factors. In the case of the present manuscript, this knowledge is then applied to the context of HCI. Hence, although the target readership of the present thesis is mostly the HCI community, the results developed here might be of interest to other communities studying motor control, especially Chapter 6.

Organisation of the Thesis

The manuscript consist of three parts. There is no general Appendix at the end of the manuscript; instead an Appendix follows each chapter, in the hope of increasing the manuscript's readability.

Part I consists of three chapters, and builds mostly on existing results from the information-theoretic, motor control and HCI literatures.

The first chapter is a historical perspective of Fitts' law needed to make sense of the unusual combination of information theory, motor control and HCI. I will also use this chapter to introduce a number of concepts of Shannon's information theory that will be used throughout the thesis.

The second chapter describes known regularities of human aimed movement. A successful model should embrace the variability inherent to human movement, be able to account for feedforward as well as feedback control, and should treat intermittent as well as continuous control. This chapter is essential to capture the diversity that makes aimed movements challenging to apprehend and model.

The third chapter discusses methods and controversies around the use of Fitts' law in HCI, which the thesis hopes to address, partially or fully. The many formulations of Fitts' law are discussed, as is the issue of considering nominal or effective quantities when measuring human performance. Finally, the all-encompassing notion of throughput is considered.

Part II contains the main results of this thesis, through three models. The first model is a simple source coding model, i.e. a model of transmission without noise.

I then take advantage of this model in the next chapter to provide a complete transmission model (FITTS 1), whose purpose is to formalize Fitts' work. Fitts' law for the error-less case is derived by computing the capacity of a channel with a uniformly distributed and amplitude limited noise. I then show that the uniform noise provides a useful bound for any noise distribution, and later extend the model to the case where pointing errors are permissible. In doing so, I propose a new index of difficulty $\mathrm{ID}(\varepsilon)$. Finally, the consequences of using a capacity result are interpreted in the context of movements: only the "best" movements should be considered when using Fitts' law, through the so-called front of performance [66, 67].

The final model (FITTS 2) is the crux of the thesis. Building on the regularities of human movement identified in the previous part, I propose a model which describes whole trajectories, accounts for feedback information and can deal with intermittent or continuous control. I propose the study of Positional Variance Profiles (PVP), and show how they can be exploited to describe human movement into two phases: a first phase where variance increases over time and where most of the distance to the target is covered, followed by a second phase where the variance decreases until the task's constraints are satisfied. I show that when the participant is performing at full capabilities, the PVP profile during the second phase is expected to decrease exponentially over time at a rate which can be interpreted as the Shannon capacity of the human.

Part III puts the models from Part II in a more practical context. FITTS 1 suggests the new concept of front of performance, for which empirical methods need to be worked out. I show how an exGauss distribution can meet the constraints of the front of performance while acknowledging that most trials are performed without the participant's full implication, and show how Maximum Likelihood Estimation can be used in this context. I show that the exGauss distribution can successfully model very diverse pointing data, from laboratory to field study datasets.

FITTS 2 entails a new approach to pointing and a number of hypotheses need to be investigated. I show that the first phase has a constant duration in Fitts' stereotyped paradigm and validate the exponential rate of decrease of the variance during the second phase

A conclusion finishes this work and provides new perspectives for HCI, through the study of atypical populations and a finer characterization of input techniques.

Notations

- $\log$ is the natural (base e) logarithm.
- $\log_X$ is the base X logarithm ($\log_X(x) = \log(x)/\log(X)$)
- $\mathbb{E}$ is the mathematical expectation.
- $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution with mean μ and variance σ^2.
- $p_X(x)$ is the probability density function of the random variable X.
- $p_{X|Y}(x|y)$ is the probability density function of the random variable X conditional on Y.
- $H(X)$ is the entropy of the random variable X, computed from its pmf or pdf $p_X(x)$.
- $I(X;Y)$ is the mutual information between X and Y, computed from their joint pmf or pdf $p_{XY}(x, y)$.
- X^n is the list notation $X^n = \{X_1, X_2, \ldots, X_i, \ldots, X_n\}$.
- erf is the Gaussian error function, equivalent to the z-score and to Marcum's Q-function.
- $Q(F)$ is the quantile function, defined as the inverse of the CDF.

Abbreviations

- MT: Movement Time
- ID: Index of Difficulty
- HCI: Human Computer Interaction
- pmf: probability mass function
- pdf: probability density function
- SNR: Signal to Noise power Ratio
- CNS: Central Nervous System
- LSE: Least Squares Estimate
- OLS: Ordinary Least Squares
- AIC: Akaike's Information Criterion
- MLE: Maximum Likelihood Estimation
- RSS: Residual Sum of Squares
- MMSE: Minimum Mean Squared Error
- PVP: Positional Variance Profile
- GEV: Generalized Extreme Value
- CDF: Cumulative Density Function
- QQ-plot: Quantile-Quantile-plot

Contents

List of Figures

List of Tables

Part I

Preliminaries

Chapter 1

Shannon's Information Theory, Psychology, and HCI

Fitts' law, widely used in HCI, was derived from information-theoretic principles. So how did information theory —or communication theory as Shannon liked to call it— come to play a part in HCI ? It is indeed surprising that this framework, developed mostly in an effort to increase the efficiency of communication systems (e.g. data compression schemes, with the ZIP, MP3 and JPEG formats, channel error correction schemes used for DSL lines, CDs) can prove useful for modeling human performance and evaluating input devices.

One could think that some HCI researchers with a computer science background might have heard of Shannon during their curriculum, and that they would have introduced his results to the field. This is far from what actually happened. As we will see in this chapter, there was a time, albeit short, when mathematicians, engineers, philosophers, psychologists, biologists, anthropologists [79, 130] etc., gathered to discuss the *human* from a systemic point of view. This was the opportunity for a theory such as Shannon's to expand well beyond his community, and towards, for example, experimental psychologists. It is then through the human factors applied to computing that Shannon was brought into HCI.

This chapter recounts this period, explains the genesis of Fitts' law, and introduces useful results that will be used throughout the thesis. I believe that this historical approach is necessary to capture what initially motivated this work — a sort of game, to find how close we could get to a formal derivation of Fitts' law using Fitts' original ingredients i.e., Shannon's capacity formula for the Gaussian channel and Fitts' pointing paradigm.

1.1 Post-war Cybernetics

A post-war scientific movement, known as *Cybernetics*[1], sought to unify several fields of science through the notion of *information*. It is because of cybernetics that words such as input, output, codes are part of the psychologist's vocabulary, and that words such as bandwidth and capacity are commonly used in HCI. There are many accounts of the cybernetics group and the effervescence surrounding it, and the goal here is not to explain in details the whole period. Instead, we try give some clues to understand how mathematicians came to collaborate with psychologists, starting with Wiener's account of the design of an anti-aircraft cannon.

1.1.1 Wiener's Anti-Aircraft Cannon

In his autobiography, Norbert Wiener explains that at the beginning of the war, many researchers and engineers were looking " in what sector of work [they] might make [them]selves of use" [162].[2]

During the war, Wiener worked on improving the design of anti-aircraft cannons, which were tricky to operate effectively. An operator of an anti-aircraft cannon must aim ahead of the plane he is trying to shoot down. The operator is thus faced with a complex task of prediction as the pilot will usually engage in audacious maneuvers to escape the ballistic shell. The design of a "smart" cannon is conditioned by the limits of human motor performance; the plane is controlled by a pilot who can only react so fast and only withstand so many dynamical constraints, and the operator who controls the cannon faces similar constraints. Providing an adequate model for the human and its extreme performance thus proved essential, and collaboration between behavioral psychologists and engineers was therefore required —although Wiener would ultimately propose a solution that would remove the human completely out of the loop.

Wiener adapted results from integral equations he had developed some time before to solve the prediction problem; this solution was then used to control the cannon using feedback to avoid modeling too precisely the cannon itself: " Not only is a feedback system less dependent on changes of load than a system without feedback, but this dependence becomes increasingly less as more and more motion is fed back" [162].

Wiener noticed that when the feedback was too strong, his system would start to oscillate wildly and uncontrollably i.e., would become unstable. Wiener made a parallel between human controlled movement and servo-mechanisms, which already implemented feedback at the time and asked a neurophysiologist, Dr Rosenblueth, whether humans could also display signs of instability. Rosenblueth informed him of *intention tremors*; wild uncontrolled oscillations that would perturb simple tasks, such as grasping a glass of water, but would be unapparent at rest. From this point, Wiener would

[1] Cybernetics is still represented today, through e.g. IEEE's Systems, Man and Cybernetics (SMC) society.

[2] His quest would ultimately set the premise of a complete theory of prediction and filtering, and would plant the seeds of cybernetics.

see the human nervous systems as a complex system made up of multiple feedback loops; and he would eventually analyze many other fields such as economics, politics, sociology with the same scrutiny [161, 163].

1.1.2 Macy Conferences and Cybernetics

Six months after the end of World War II, a set of conferences was initiated by the Macy Foundation that came to be known as the Macy Conferences with the following goal: [3]

> To make best use of our time, we will ask that the themes be discussed so as to make known theoretical and practical developments in the domains of computing machines and apparatuses that aim for their targets, and in each case exemplifications in physiology should be presented. After, we will consider problems of psychosomatic, psychological, psychiatric and sociological nature where these notions are applicable and we will see which extensions of the theories are necessary for these issues.

These conferences would gather, among others, Wiener (mathematician, "father of cybernetics", known for his work on Fourier transforms and Brownian motion), Von Neumann (mathematician)[4] , Shannon (mathematician, "father of information theory", also known for applying Boolean logic to electrical circuits design and work on cryptography), Bar-Hillel (mathematician and linguist), Bateson (linguist and anthropologist), Bigelow (engineer who had worked with Wiener on the anti-aircraft gun), Hutchinson ("father of modern ecology"), Lewin (psychologist), Licklider (psychologist and computer scientist), Luce (mathematical psychologist), MacKay (physicist, known for his work on the theory of brain organization), McCulloch (neurophysiologist), Morgenstern (economist), Pitts (logician), Quastler (radiologist), Wiesner (electrical engineer), Young (zoologist and neurophysiologist). A sample of the topics discussed throughout the Macy conferences (1946–1953) illustrates the diversity of the talks delivered:

- Self-regulating mechanisms,
- Anthropology and how computers might learn,
- Perceptual effects of brain damage,
- Analog versus digital approaches to psychological models,
- Memory,

[3] The following excerpt is a translation to English from a French version given by Segal [130]. However, that French version is in itself a translation from an English version; the original source is unfortunately hard to find.

[4] Von Neumann's contributions are hard to compile, so many have they been; Von Neumann's biograph summarizes that:"If the influence of a scientist is interpreted broadly enough to include impact on fields beyond science proper, then John von Neumann was probably the most influential mathematician who ever lived" [123]

- Deriving ethics from science,
- Compulsive repetitive behavior,
- The applicability of game theory to psychic motivations,
- Collaboration between physics and psychology,
- Analog versus digital interpretations of the mind,
- Language and Shannon's information theory.

These conferences established the fundamental nature of feedback "information"[5] as well as the usefulness of the notion of *message*. The Macy conferences were also an opportunity to show that some engineering and mathematical concepts had a general significance well beyond technical points, and could lead to interdisciplinary synthesis, particularly around the notion of *information*, and could prove useful in e.g., the social sciences.

1.1.3 Annus Mirabilis: 1948

In 1948, Claude Shannon published *A Mathematical Theory of Communication* [132], a paper that pioneered the modern analysis of digital communications and established *entropy* as a relevant measure of information.

Shannon's theory supposes a generic paradigm for communications, as represented in Fig. 1.1, composed of a *source*, an *encoder*, a *channel*, a *decoder*, and a *destination*. The source is modeled as a random[6] process, and both the encoder and the decoder are expected to know and take full advantage of the statistical properties of the source and the channel.

- The information *source* produces a message, modeled as a random variable, say M, which takes values in a set $\mathcal{M}$. The only aspect that matters is that we can assign a probability to each outcome: "Semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one selected from a set of possible messages" [132]. The average amount of information produced by this source, when it emits a message, can be quantified using the entropy H. For a discrete random variable M, with a probability mass function (pmf) $\mathbf{p}(M)$,

$$H(\mathbf{p}(M)) = -\sum_{M\in\mathcal{M}} \mathbf{p}(M)\log_2 \mathbf{p}(M) = -\mathbb{E}\log_2 \mathbf{p}(M) \text{ bits.} \tag{1.1}$$

[5] The concept of information was loosely applied at first. Quickly, several mathematical definitions, such as Shannon's [132], measured in bits, or Gabor's [46], measured in logons, were proposed.

[6] In the sense that one cannot know with certainty the output of the source, even knowing the previous output.

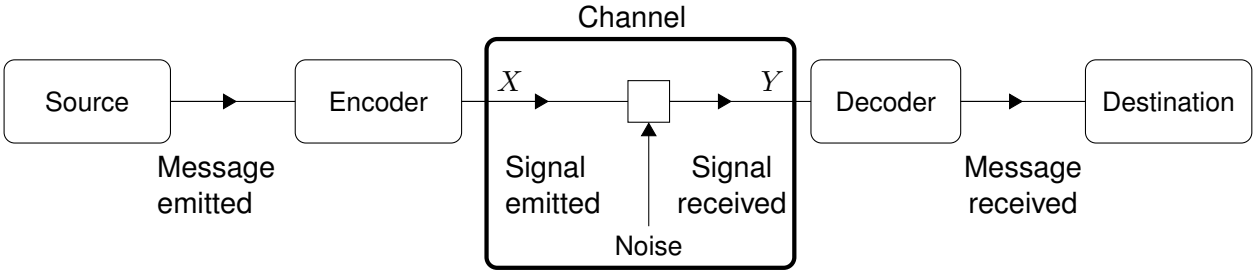

Figure 1.1: Shannon's point-to-point communication paradigm

Notice that entropy is not a function of M, but rather of its associated pmf $\mathbf{p}(M)$ and therefore does not depend on the values M assumes; however as is usually accepted, we will lighten the notation by writing $H(M)$ rather than $H(\mathbf{p}(M))$. Further notice that a similar quantity can be obtained for a continuous random variable by replacing the sum by an integral and the pmf by a probability density function (pdf); the formulation that uses the mathematical expectancy $\mathbb{E}$ conveniently generalizes both cases and will therefore be mostly used throughout the manuscript. An instrumental quantity that takes into account the dependency between messages, known as the rate of information R can be obtained by dividing the entropy of a sequence of messages (symbols) by the number of symbols n of the sequence, or the time T needed to produce this sequence:

$$R = \frac{1}{n} H(X_1, X_2, \ldots, X_n) = \frac{1}{n} \mathbb{E} \log_2 \mathbf{p}(X_1, X_2, \ldots, X_n) \text{ bits/symbol, or} \quad (1.2)$$

$$R' = \frac{1}{T} H(X_1, X_2, \ldots, X_n) = \frac{1}{T} \mathbb{E} \log_2 \mathbf{p}(X_1, X_2, \ldots, X_n) \text{ bits/seconds.} \quad (1.3)$$

The rate represents the minimal bit rate at which it is possible to encode the sequence of n messages without compromising its quality. It is thus useful when one looks to compress a source i.e., to convey the same amount of information with fewer symbols. When there is no inter-symbol dependency i.e., the messages are independent from each other and are identically distributed, then

$$R = \frac{1}{n} H(X_1, X_2, \ldots, X_n) = \frac{1}{n} \sum_{k=1}^{n} H(X_k) \quad (1.4)$$

$$= \frac{1}{n} \times nH(X) = H(X), \quad (1.5)$$

and the rate reduces to the entropy.

- The *encoder* adapts the message from the source to the channel, in at least two aspects: a physical adaptation in which the message is converted into a suitable signal for transmission (e.g., the variation of an electrical current); and a channel encoding in which certain operations are performed on the message to

enhance transmission quality. The important feature is that the encoder performs *deterministic* operations and is thus nothing more than a function —albeit usually this function is complex and operates on long sequences of symbols— say f, and the output X of the decoder is given by

$$X = f(M). \tag{1.6}$$

- The *channel* is the medium that serves to transmit the signal from the *emitter* (source and encoder pair) to the *receiver* (decoder and destination pair). On its way from the emitter to the receiver, the signal may be corrupted by *noise*, so that the output of the channel does not perfectly correspond to its input anymore. Some dependence of the output on the input can subsist —this is actually why reliable transmission is possible— and can be estimated through mutual information. If X is the input to the channel and Y its output, then the mutual information $I(X;Y)$ is given in the following equivalent forms:

$$I(X;Y) = \mathbb{E}\log_2 \frac{\mathbf{p}(X,Y)}{\mathbf{p}(X)\mathbf{p}(Y)} = H(X) + H(Y) - H(X,Y) \tag{1.7}$$

$$= \mathbb{E}\log_2 \frac{\mathbf{p}(Y|X)}{\mathbf{p}(Y)} = H(Y) - H(Y|X) \tag{1.8}$$

$$= \mathbb{E}\log_2 \frac{\mathbf{p}(X|Y)}{\mathbf{p}(X)} = H(X) - H(X|Y) \text{ bits}, \tag{1.9}$$

where $\mathbf{p}(X,Y)$ is the joint pmf (or pdf) of X and Y, and $\mathbf{p}(Y|X)$ is the conditional pmf (or pdf) of Y knowing X — incidentally, knowledge of $\mathbf{p}(Y|X)$ fully determines the channel. Mutual information quantifies the elusive notion of "transmitted information", through a measure of the statistical dependency between two categorical variables (Eq. (1.7)), and equivalently, as the difference between the receiver's uncertainty about X before the transmission $H(X)$ and after the transmission (i.e., given the channel output) $H(X|Y)$ (Eq. (1.9)).

- The *decoder* performs deterministic operations to get back to the message space from the signal space. If Y is the output of the channel and $\widehat{M}$ the received message, then

$$\widehat{M} = g(Y). \tag{1.10}$$

If the decoded message differs from the message that was sent by the source i.e., $M \neq \widehat{M}$, we say there is an error; we can give a simple measure of the reliability of the transmission scheme through the probability of error after transmission P_e:

$$P_e = \mathbb{P}(M \neq \widehat{M}). \tag{1.11}$$

As they result from the environment, channel errors are inevitable; however one of Shannon's most fundamental result expresses that, surprisingly, a system can be

designed such that errors *at the receiver* vanish. This result is known as the channel coding theorem, in which mutual information plays an instrumental role, and is probably one of Shannon's most useful, operational result.

Theorem (Channel Coding Theorem)**.** *The capacity of a memoryless channel*

$$C = \max_{\mathbf{p}(X)} I(X;Y),$$

expressed in bits per channel use (bpcu), is such that for any rate $R < C$ and any $\varepsilon > 0$, there exists a coding scheme leading to an arbitrarily small probability of error $P_e < \varepsilon$.

Usually, there are physical limitations on the input, so that C is computed from maximizing mutual information over all distributions for X subject to the relevant constraint e.g., the power constraint $\mathbb{E}\big[X^2\big] \leq P$. The theorem expresses that as long as the rate R does not exceed the limit C (which is found by solving an optimization problem) the error probability P_e can be made as small as we like—this defines *reliable* communication as a mathematical limit. It is noteworthy that the theorem describes a tradeoff between the rate (speed) and the reliability (accuracy) of the transmission, expressed in the case of arbitrarily low probability of errors, which suggests there is reason to believe in an information-theoretic model for the speed-accuracy tradeoff.

This theorem, when applied to the case of the Gaussian channel with an input power constraint, produces what is probably the most famous result from Shannon's paper [132]:

Theorem (Shannon's Theorem 17 [132])**.** *The capacity of a channel of band B_W, perturbed by white thermal noise of power N, when the average transmitter power at the input is limited to P, is given by*

$$C = B_W \log_2 \frac{P+N}{N} \tag{1.12}$$

$$= B_W \log_2\big(1 + \textit{SNR}\big) \textit{ bits per seconds,} \tag{1.13}$$

where SNR $= P/N$ is the Signal to Noise power Ratio.

As we will see in the next section, several of Shannon's ideas have been applied by researchers well beyond engineering. Entropy, as a measure of uncertainty, as well as Shannon's capacity formula for the Gaussian channel were particularly successful among e.g., psychologists. This is partly due to the post-war Macy conferences; it is also without a doubt largely due to the re-edition of Shannon's seminal paper of 1948 a year later in a book format, preceded by a noteworthy introduction by Weaver [134]. Rioul [135], in the preface to the most recent French translation of the book, explains the following:[7]

[7] This is a personal translation from French to English; emphasis added.

> Weaver takes a stance opposite to that of Shannon: while the latter excludes the semantic aspects of messages from the outset to tackle the technical problems faced by the communication engineer, Weaver seeks to explain how Shannon's ideas could be extended far beyond its initial reach, towards all sciences that handle communication issues in the broadest sense – such as linguistics and social sciences. Warren Weaver's ideas have had a tremendous impact, precisely because they appear *before* Shannon's text in the original version of the book: it is likely that many readers produced their opinion of the theory by reading Weaver and stopped reading after Shannon's first mathematical results. Today, the theory is sometimes attributed to *both* Shannon and Weaver, and the first French edition of this book by the Retz editions in 1975 even credits Weaver as the first author.

At the same time, Wiener published a book that would have an even greater impact on the layman, *Cybernetics: Or Control and Communication in the Animal and the Machine*, which essentially popularized some of the discussions at the Macy Conferences about information, feedback and the digital computer, as well as their applications to living systems. Although it contains many technical aspects (including pages full of equations), the book aroused substantial public interest, possibly because Wiener addressed not only engineering problems but political, philosophical, and social issues, sometimes with a prophetic stance. In an American magazine, *Saturday Review*, Wiener's book is described as a "must, for those in every branch of science–engineers (all kinds), mathematicians, physiologists, psychologists, sociologists, anthropologists, chemists (all kinds), psychopathologists, neuroanatomists, neurophysiologists etc. In addition, economists, politicians, statesmen, and businessmen cannot afford to overlook cybernetics and its tremendous, even *terrifying*[8], implications." [145]. In the same magazine, Wiener's work is included in a list of books, targeted at an auditory of businessmen, dealing with economics, management, and labor [108].

1.2 Psychologists Discover Information Theory

1.2.1 Miller and Frick's Statistical Behavioristics

Influenced by the concepts introduced by Wiener, Von Neumann and most importantly Shannon, George Miller and Frederick Frick [106] proposed a formal analysis of serial dependencies, which used many ingredients of Shannon's information theory. Serial dependencies occur when a response is not only triggered by a stimulus but is also affected by a preceding response, and are common in psychophysics experiments. Pinker [116] explains that Shannon's seminal paper had come as a revelation to Miller; and Miller and Frick indeed presented a statistical account of *sequential* responses using state diagrams, joint probability distributions, conditional probability distributions, probability trees, a logarithmic measure of uncertainty and Markov processes, all of

[8] The emphasis is in the original review.

which were described in Shannon's paper. By translating Shannon's work into the language of behavioral psychology, Miller and Frick had just presented psychologists with a "shiny new tool kit, and a somewhat esoteric new vocabulary to go with it" [3]. This was in 1949. Ten years later, in 1959, Attneave [3] would ascertain that "more than a few psychologists had reacted with an excess of enthusiasm" and that "although some of the attempts to apply informational techniques to psychological problems were successful and illuminating, some were pointless, and some were downright bizarre".

One of the earliest successful applications of information theory to psychology was Hick's law [74], later extended by Hyman [77]. Hick's law relates what is known as the b-reaction time "the time between a stimulus and the choice by the participant of an appropriate response", with the stimulus entropy. It was found by Hick that for a choice reaction task, where the stimulus was as simple as a bulb lighting up, and the response as simple as pressing a key immeditaley below the right light-bulb, the b-reaction time increased linearly with the entropy of the set of stimuli. This experiment would be re-enacted multiple times in different contexts and with different tasks [159]; the interpretation being that the rate of gain of information was approximately constant with respect to time.

Perhaps the most memorable application of information theory to psychology is Miller's so-called magical number 7 [105]. In this highly-cited and most-influential paper, Miller observes the coincidence that absolute judgment and short-term memory share a same limit—the magic number seven (plus or minus two). In an absolute judgment task, a participant is presented with a stimulus that varies on one dimension, and learns the appropriate response to each stimuli. For example, the participant might be subjected to 10 tones of different loudnesses, to which he should respond with the corresponding number between 1 and 10. It turns out that in such tasks, performance is usually perfect up to about 5 or 6 different stimuli; the same observation holds for e.g. pitch, saline concentration, hue, brightness, vibration intensity and duration and square size [105]. Importantly, the mutual information between the stimulus and the response was found to level off as stimulus entropy increased; Miller interpreted this limit as an upper bound on transmission i.e., a channel capacity that "imposes severe limitations on the amount of information that we [humans] are able to receive, process and remember". Many other cases of the information-theoretic approach to psychology can be found in [3] and [91], and most of these model the human as an information transmission channel with limited capacity.

1.2.2 Fitts' Law

Inspired by Hick's analysis of the choice reaction time, which showed that reaction time correlates strongly with the entropy of the stimulus, Paul Fitts set out to investigate whether selecting a continuous quantity, such as movement amplitude or force would lead to similar results. Fitts presented a study focused on amplitude selection during a

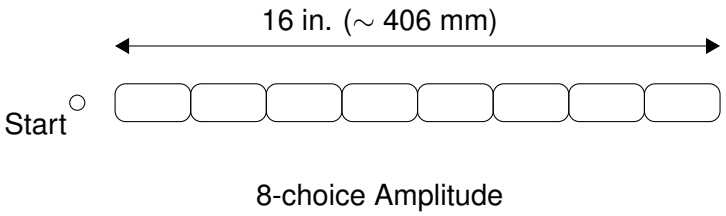

Figure 1.2: Stimulus in Fitts' 1953 [37] amplitude choice task.

lecture given in 1953 [37].[9] The stimulus he proposed to manipulate it consisted of a set of aligned rectangles, one of which was indicated to be the target, see Fig. 1.2.

Whereas in Hick's experimental design, movement time was extremely short approximately constant across conditions, this was certainly not the case with Fitts' design of Fig. 1.2, as the targets located further away inevitably lead to longer movement times. To counterbalance this effect, Fitts found out that if the targets that were farther away from the starting point where appropriately enlarged, movement time would remain constant. This result on movement time would ultimately prove to be the most important, far more than any findings he had on the reaction time.

Fitts formalized this finding by conducting the so-called tapping experiment, see Fig 1.3. The task consisted of alternatively tapping two rectangular plates. These plates had a width W, measured in the movement's direction, and were long enough (6 in., ~ 15 cm) in the orthogonal direction so that they could be considered unconstrained in height. The distance between the center of both plates D could also be varied. According to Fitts, the model that would best approximate this task is that of the *information source* [37], whereas the subjects' performance could be taken as a measure of his *capacity* for "performing repetitive motor tasks under varying conditions of response coding". As detailed in Chapter 4, this view leads to the definition of an Index of Difficulty (ID)

$$\mathrm{ID} = \log_2 \frac{2\mathrm{D}}{\mathrm{W}} \text{ bits}, \tag{1.14}$$

to which, Fitts hypothesized, movement time should be proportional.

In 1954, in what is now a seminal paper [38], Fitts presented the main empirical results presented the year before, but with a significant shift of their interpretation. Whereas in the 1953 lecture, Fitts' justification of the ID was based on an information source argument, in the 1954 paper he invoked *transmitted information* through an analogy between W and the concept of noise. He also referred to Shannon's capacity formula in a footnote yet he surprisingly continued to use the source argument to compute ID. In 1964, Fitts [40] completed the shift towards transmission, and derived ID (1.14) from a direct analogy with Shannon's Theorem 17. He reasoned that:

[9] The proceedings of which, composed of 7 lectures, have been published in the series *Current trends in psychology* and entitled *Current trends in information theory*. This once again underlines the proximity between information theory and psychology at that time.

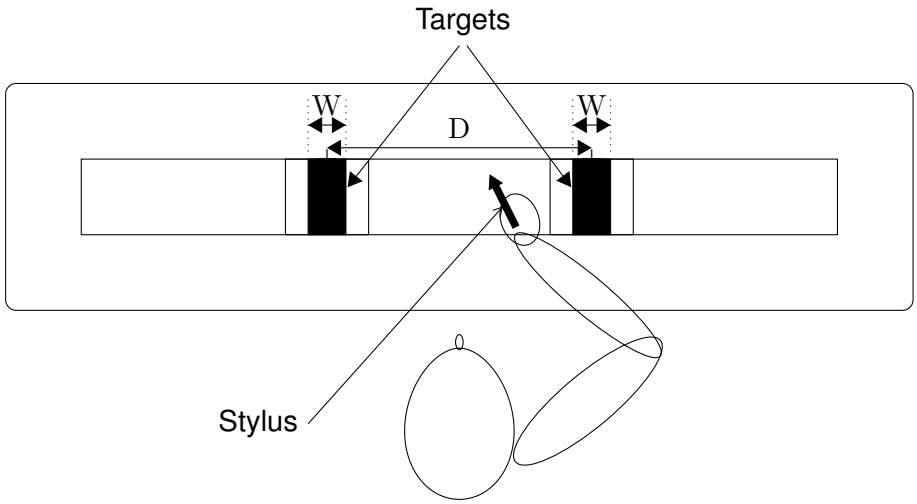

Figure 1.3: Fitts' tapping task. The participant is facing a board with two targets, whose width W can be varied. The distance separating the two targets D can also be modified. The task it to hit the center plate (target) in each group alternatively without touching either side plate (error).

- the average amplitude D of the participant's movement is equivalent to signal plus noise amplitude;
- half the range of movement variability is equivalent to peak noise amplitude.

Empirical results showed that, as long as W and D were in appropriate ranges, the ratio

$$\frac{\mathrm{ID}}{\mathrm{MT}} = C \tag{1.15}$$

remained constant, where MT is the movement time measured in the experiment, and C is man's capacity for executing a particular class of motor responses, in bits per second. With the addition of an intercept a for better fitting power, Fitts' law predicts movement time in the tapping paradigm:

$$\mathrm{MT} = a + 1/C\ \mathrm{ID} = a + b\ \mathrm{ID} \tag{1.16}$$

$$= a + b\ \log_2(2\mathrm{D}/\mathrm{W}), \tag{1.17}$$

where the *intercept* a and the *slope* b have to be estimated empirically. This formulation has evolved over time, and as discussed in Chapter 3, there exist about a dozen different expressions for ID [117]. However, the basic idea that movement time scales logarithmically with the relative distance D/W has mostly remained unchallenged.

1.3 Whatever Happened to Information Theory in Psychology ?

Today it is not uncommon to find information-theoretic approaches in statistics, probability, economics, biology [18]; however it is less so in psychology. Information theory had become so popular in the nineteen fifties that many psychologists had perhaps become over-eager to use it: Many resulting applications, which Attneave [3] had described as "pointless", or "downright bizarre", were far fetched and unfruitful.

1.3.1 Reaction from the Communication Engineers

The use of information theory outside the sphere of communication engineering was challenged by the information theory community, and Shannon himself. In a famous editorial, Shannon [133] expressed the view that information theory "has perhaps been ballooned to an importance beyond its actual accomplishments." He also insisted that "the use of a few exciting words like *information, entropy, redundancy*, do not solve all our problems." Tribus [90, p. 1] reports a private conversation he had with Shannon who "made it quite clear that he considered applications of his work to problems outside of communication theory to be suspect and he did not attach fundamental significance to them".

Elias [30], an important figure of the information theory society to which we will return in Chapter 6, urged authors—using a very ironic, even aggressive tone—to stop writing approximative papers that abused information theoretic results and concepts. Fitts' work, based on a loose analogy with Shannon's Theorem 17, is a good example of abuse of information theory:

- Why should D/W of Fitts' law be analogous to P/N as defined in Shannon's Theorem 17?
- What is the bandwidth B_W of Shannon's Theorem 17 analogous to in Fitts' law ? There seems to be no reason to identify B_W to $1/\mathrm{MT}$ beyond the fact that both are expressed in the same physical units (s^{-1}).
- Since D and W are amplitudes while P and N are powers, what happened to the squares[10] ?

More importantly, Fitts never made explicit any communication scheme for the aiming task; no serious information analysis can dispense with such a scheme. Whether Fitts attributes the law to a limit in *transmission* due to the channel, or to a limit in *generation* due to the source of information is also unclear.

[10]The power P of a random variable, say X, is the average of its squared values $P = \mathbb{E}[X^2]$. If X is a centered i.e., zero-mean random variable, then power is equivalent to the variance.

1.3.2 Reactions from Psychologists

Criticism also emerged from within the psychological community, particularly from those who had the necessary mathematical background to fully understand Shannon's theory. Luce [91], a prominent mathematical psychologist characterizes the relationship between information theory and psychology as follows:

> In the attempt to analyze communication systems, a mathematical formalism has been produced [...] and this mathematics can be completely divorced from its realization as a communication system. At the same time, there are other realizations of the same mathematical system in psychology.

Indeed, concepts like entropy or mutual information allow very simple characterization of sequential responses, especially when expressed on a categorical, ordinal or interval scale and there is no proper metric to quantify the relevant quantity. These concepts produce results that are comparable across many different experiments, as the final analysis is inevitably expressed in bits or bits per second. The fact that information theory can be used in psychology has thus nothing to do with the semantics of information, or the concepts of capacities, but more to its usefulness as a mathematical tool. In fact, almost none of the applications to psychology *really* use the fundamental theorem relating channel capacity, the statistical structure of the source, and the transmission rate [91].

In practice, what information theory actually brings to the table for psychologists is quite limited; Miller [104] remarks that "information theory effectively corresponds to a nonparametric analysis of variance", with two inherent difficulties in empirical psychological studies [91]:

- Sequential responses obtained while the participant is going through a learning phase do not allow a reliable estimation of the probabilities that are needed: learning is adverse to stationarity needed for estimation,
- As dependencies between stimuli extend and numbers of dimensions increase, the number of samples that are needed completely grows out of hand.

Mutual information is not the only measure that can describe the relationship between two variables, yet comes with serious empirical difficulties. In addition, it effects a particular compression of the data [19], which might not always be warranted. For instance, it completely neglects constant errors, as in the communication scheme these are supposed to be known at the receiver, who can then easily correct them. In many cases, limits described as channel capacities could be ascribed to some reasonable and known physiological limitation [70, 122], and the psychologist may be better off with a different framework.

1.3.3 Fitts' Law is Not Information-Theoretic Anymore

In retrospect, Attneave's survey of 1959 looks like a funeral tribute. Since the end of the sixties very few new articles in psychology have referred to information-theoretic

principles. Fitts' derivation of the ID and his reasoning by analogies is a good example of the type of work that would have bothered Shannon and consorts.

In 1963, Crossman and Goodeve [22] proposed a novel explanation for Fitts' law that did not rely on information-theoretic results. Their model, based on feedback considerations, assumed an aimed movement to be composed of a sequence of sub-movements each of fixed duration and covering a fixed fraction of the remaining distance. The logarithmic nature of the law was not attributed to man's limited ability to transmit information anymore, but to a visual and/or kinesthetic iterative feedback mechanism. Although the model provided a nice rationale, it faced a number of limitations, mostly because it was deterministic: it failed to explain movement end-point variability and excluded the very possibility of target misses.

By the end of the eighties, Meyer et al. [102, 103] proposed a *stochastic* feedback mechanism for rapid aimed movements, thus eliminating the main flaw of the Crossman and Goodeve model. Meyer et al. proposed what they called a power model of Fitts' law, rather than a logarithmic one. In fact, as shown by Rioul and Guiard [124], mathematically the Meyer et al. model falls in the class of quasi-logarithmic models. The stochastic optimized sub-movement model of Meyer et al. [102] is now considered by many psychologists (e.g. a reference textbook on motor control [127]) as the leading explanatory theory of Fitts' law, illustrating the extent to which the information theoretic approach has lost ground in modern experimental psychology.

1.4 Fitts' Law, Shannon's Theory, and Human Computer Interaction

Although Fitts' law is rarely addressed through Shannon's framework anymore in psychology, in Human Computer Interaction the information theoretic explanation of the law has survived up to present.

1.4.1 Bringing Fitts' Law in HCI

HCI emerged as a specialty in computer science that embodied human factors and cognitive science, to provide better access to computers through a more efficient experience for the user. In 1978, a seminal paper by Card et al. [11] introduced Fitts' law to evaluate the performance of desktop input devices.

One of the hypotheses explored by the paper was whether Fitts' law could describe how humans moved a cursor around the screen to reach a target. It was unknown whether results obtained using Fitts' simplistic task would be comparable to those obtained when selecting real-life common objects on computer displays such as text strings, with an input device.

The performance of four input devices, among which the mouse, was evaluated by measuring the time required to select a highlighted string from a text. The distance to the string as well as the size of the string were manipulated, thereby essentially

replicating Fitts' experiment in the context of text editing on a computer display. This study is often cited as a major factor for the commercial introduction of the mouse [138] as the mouse was shown to outperform the other devices (joystick, step keys, text keys) on both speed and accuracy. It was also apparent that Fitts' law would be a good model for selection time with a device such as the mouse, while the value of the slope b (1.17) of about 0.1 seconds per bit was typical of that found for experiments *without* input device run by experimental psychologists [158, 11]. Card et al. interpreted this as the sign that the time needed to position the cursor on a target on the screen was due to a limitation in the central information processing capacities of the eye-hand guidance system [11]. Since then, Fitts' law has been widely used in HCI to assess input devices that move a cursor, such as mice, trackballs, styluses, touchscreens, or head tilted devices [56].

Unlike experimental psychologists, HCI researchers have apparently remained confident in the promise of the information-theoretic approaches. Hick's law, although less popular than Fitts' law in the HCI community [131], is still used to model reaction time in command selection [17]. Scott MacKenzie has produced a sustained effort to develop a complete performance model of Fitts' law for HCI using the tools of information theory [93], including a modification of Fitts' formula to make the pointing analogy more consistent with Shannon's Theorem 17

$$\mathrm{MT} = a + b\ \log_2\big(1 + \mathrm{D}/\mathrm{W}\big). \tag{1.18}$$

This formulation is known in HCI as the Shannon formulation. As we will see later however, the derivation of this formulation owes little to Shannon's theory. We thus call this the MacKenzie formulation. MacKenzie [94] later incorporated information-theoretic results such as the entropy of a Gaussian distribution to account for target misses in pointing based on prior work by Crossman [21] (see Chapter 3). Importantly, the recent ISO standardization for the evaluation of pointing devices is explicitly based on information-theoretic principles [142, 1].

1.4.2 Importance of Fitts' Law for the HCI Community

Fitts' formula (1.17) provides quantitative clues for designers and researchers to construct more efficient interfaces and techniques [4]. Its main purpose is to summarize rather complex interactions with the GUI into a single, simple formula with two parameters, the intercept a and the slope b in Eq. (1.17). These parameters have two primary purposes in HCI [142].

First, once a and b have been estimated in a controlled experiment using a particular device in a specific context, the average time needed by a participant to click on a well defined target within this context and with this device can be computed directly from Fitts' law without further empirical work. Therefore, Fitts' law can be used directly to guide interface design. It also works as an elementary block to build complex models

of interaction, especially in the family of Goals, Operators, Method and Selection rules (GOMS) models [10] and extensions thereof (e.g. StEM [51]).

The second purpose of Fitts' law in HCI is to act as a comparison tool between different interaction techniques and different contexts, through the so-called throughput. As will be explained in Section 3.4, there is no unique definition for throughput [171]; however all definitions have in common that they produce the throughput measure from Fitts' law's parameters b and/or a, and they are all expressed in bits per second. A higher throughput is interpreted as the indication of a better interaction, in the sense that information is transmitted more efficiently.

There is a third reason why Fitts' law is important, which has to do with the status of psychology in HCI. In their early essay entitled *The prospects for psychological science in human-computer interaction* [112], Newell and Card discussed the role psychology should play in human-computer interaction. They describe how qualitative "soft sciences" such as psychology and human factors are gradually driven out by the more technical and/or quantitative "hard sciences". This situation, they argue, is unfortunate, as a good comprehension of the human is paramount to the design of effective interaction, and should be fought by a "hardening" of psychology. Fitts' law plays a crucial role towards this goal, as it is the formula in HCI that comes closest to a natural law such as one can find in e.g., physics, is driven by a mathematical framework and happens to come from psychology.

1.5 Discussion

In a suggestive title, "*Whatever Happened to Information Theory in Psychology?*", Luce [92] explains that information theory is "no longer much of a factor" in psychology; information theory is essentially relegated to the rank of a historical curiosity. One can thus ask why the information-theoretic rationale has remained popular in HCI.

It seems to me that the position of the HCI community is not completely honest on this issue. Fitts' law practices in HCI are justified through information theoretic concepts [142]; yet most researchers will agree that the information-theoretic analogy is precisely that — *just* an analogy.

A look at Shannon's results provides conditions *sine qua non* to use the information-theoretic tools without forcing them upon the problem. We should identify a communication scheme, whose goal is to transmit a random variable to some destination over a noisy channel. Furthermore our problem should be expressed in terms of mathematical expectations, and it should only address optimal movements i.e. those that actually reach the capacity.

Currently no model satisfies these requirements — the reasoning by analogy with information-theory has produced some results which simply *happen* to provide good fits with empirical data. Hence, the analogy is not "just an analogy", but a lucky one at that.

A recent view expressed by Laming [88] is that information theory, as a technique for investigating the behavior of complex systems, analogous to systems analysis of

physical systems, has great potential; essentially it allows the researcher to model the brain without needing to model its neural responses. Surely, these types of models are useful in HCI, where operational results are a must, and should, in principle, be possible to develop for aimed movements. At least this is the gamble we took when starting this thesis.

Chapter 2

Speed-Accuracy Tradeoff: Empirical and Theoretical Exposition

We study motor control to understand how exactly the brain directs the body to interact purposefully with its environment. The scope of this issue is evidently large, and therefore several communities exist that investigate motor control (physics, engineering, statistics, behavioral and cognitive science and human factors, physiology, neuroscience, medicine [127]), and each community brings its own set of methods to the table. The literature on voluntary movements is thus extremely vast, and it is frankly quite impossible to explore this literature exhaustively and to provide a comprehensive account of the topic. To simplify, we can consider that approaches to the study of movement come in two ways:

1. A bottom-up approach, which starts from low-level observations on the neural circuitry of the brain and wishes to understand e.g., what role the primary motor cortex exactly plays in the generation of motor commands, why there are multiple motor areas in the brain, how the cerebral cortex organizes the stream of incoming sensory information, etc. [82]. At this level, models usually deal with complex neuronal systems, but rarely translate to high level models about voluntary movement.

2. A top-down approach, where mathematical models are created to make sense of high-level observations, such as speed-accuracy tradeoffs, laws of movement (e.g., Fitts' law, Viviani's two-third power law), kinematic regularities etc. Essentially, the goal here is to reverse-engineer the human body, and find appropriate models for existing, well documented regularities. This approach rarely translates to low-level information about functioning of neural systems.

In this manuscript we adopt the latter approach; the main reason being that voluntary movements are quite complicated, far more complex than reflex or rythmic move-

ments [82]. We tend to underestimate the difficulties associated with voluntary movements, probably because they are ubiquitous. This would be forgetting that tasks that seem perfectly obvious to us now were accompanied by many struggles during learning. Voluntary movements depend on context and environment, and involve processes that are sensory, perceptual and cognitive. These processes can occur in multiple neural structures and neural populations, which makes a bottom-up approach particularly hard if one wishes to provide operational macroscopic results—precisely those of interest to an applied field such as HCI.

Therefore, we report essentially from behavioral studies and computational models. We first begin with an analysis of the regularities of voluntary movements, then follow by a description of some models relevant for our work. We also comment on their agreement with the regularities of voluntary movements.

A study of methods and controversies around Fitts' law specifically is given in the next chapter.

2.1 Characteristics of Human Voluntary Movements: An Empirical Survey

Study of invariants is a powerful process in science, and a successful model for pointing should account for invariants of human aimed movement. However, as discussed in the next three subsections, these are not easily found. Indeed, the execution of movements within the same task can differ drastically, and even the processes underlying movement are hard to capture. Presently, it is recognized that movements result from feedback and/or feedforward control and that this control may be continuous or intermittent.

2.1.1 Variability

Human movement is inherently variable [147]. If one is repeatedly given the same pointing task, both one's trajectories and summary performance measures can be strikingly different between trials. Variability affects practically all kinematic markers (e.g., position, speed, acceleration, or jerk) [38, 142, 89, 78]. Fig. 2.1 shows movement time MT in a Fitts task as a function of task difficulty ID, for a single participant of the GFA dataset (see Chapter 7). The high variability of the data, as seen by its vertical spread, is characteristic of controlled experiments with humans. It is common to have certain movement times reach twice the level of others. For example for ID levels above 4 in Fig. 2.1, the shortest movement times are about 600 ms, while many trials correspond to movement times in the interval $[1000$ ms$, 1500$ ms$]$, and two trials even lead to movement times close to 2 seconds. Variability in web-based experiments [52], or so-called "in the wild" setups is even higher [14].

One cause of the variability of pointing data is the fact that a single pointing movement can involve many joints (finger, wrists, elbows, shoulders and even the back), see [127] for a detailed account. The human body provides many redundant degrees

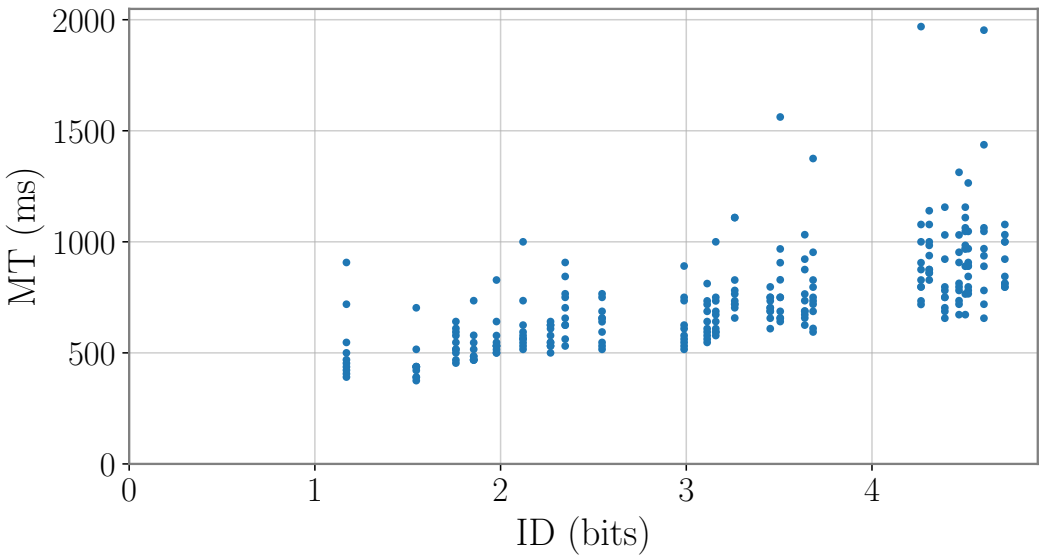

Figure 2.1: Variability in movement times MT for a Fitts task, as a function of task difficulty ID for a single participant of the GFA dataset (see Chapter 7). Each blue dot is the movement time for one trial.

of freedom that allow almost infinitely many combinations of elemental joint movements to execute the same task at the end of the kinematic chain of the arm [8]. Even though hand trajectories are far more invariant than joint trajectories [82], a chronometric analysis of the trajectories produced by a single participant aiming towards a point under the instruction to achieve maximum accuracy reveals significant positional variability (Fig. 2.2) of the hand. Although all movements start and stop at identical locations, the trajectories in between display significant spatial and temporal variability.

Other sources of variability are the noise in the planning and execution of movements [148] as: “no part of the system, from the neurons firing in the brain all the way to the muscle motor units is of a deterministic nature” [150]. Indeed, the variability is intrinsic to our sensors [100] and motor neurons because of fluctuations in their membrane potentials [82]. This type of consideration has led many authors to treat variability as the equivalent of a noisy signal disrupting e.g. a deterministic controller or a transmitting system [100, 150, 72, 137], and even applying signal processing techniques such as spectral analysis and denoising filters to study movement [152, 151].

A third reason comes from the well-known fact that participants in an experiment are rarely fully committed to repetitive, boring tasks such as those composing e.g. a Fitts’ law experiment. It is unreasonable to take it for granted that participants routinely perform to their fullest capabilities during the course of an experiment, even if they are explicitly instructed to do so [57, 67].

Finally, in computational models, variability can also originate from the failure to take

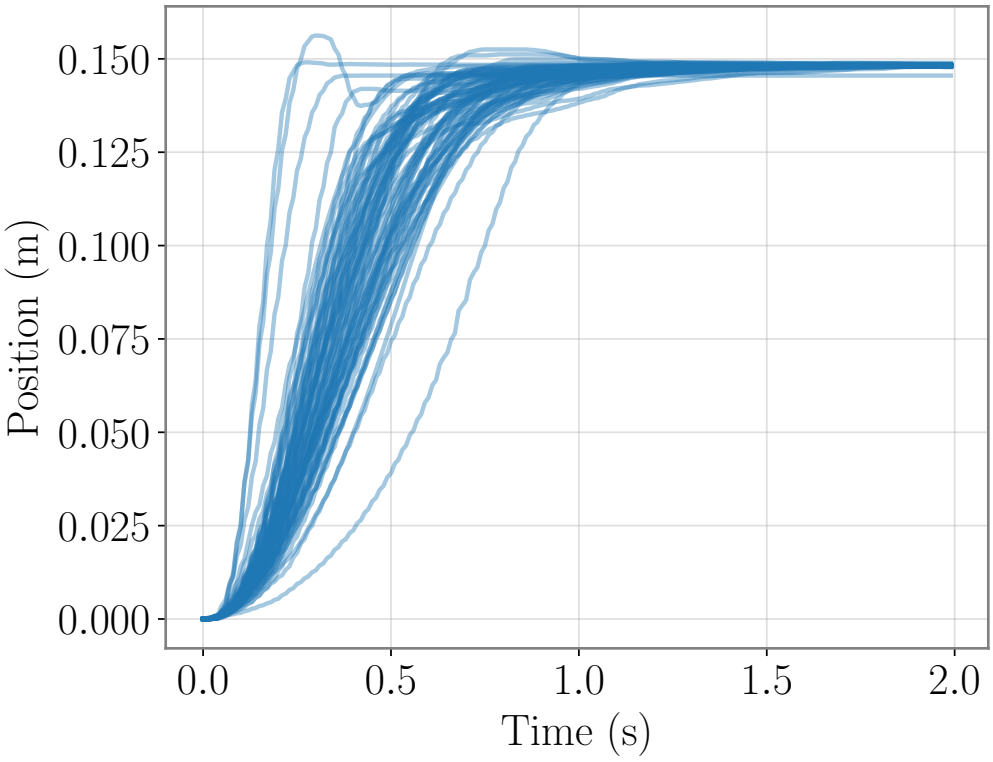

Figure 2.2: Trajectories (position as a function of time) produced by a single participant of the G-dataset (see Chapter 7) aiming towards a point, located 15 cm away from the starting point. Each blue curve is a single trajectory; The acquisition was performed with a 1-D task.

into account the different complexities of the system [146], such as non linearity [31]; it is then introduced somewhat deliberately to facilitate modeling.

2.1.2 Feedback and Feedforward Control

The body continuously monitors its state (intrinsic information) and its environment (extrinsic information) through a variety of sensors. For example, during movement, information about hand position can be given by the eye — so-called visual feedback. Muscle length and muscle contraction rate, which pertain to kinematics, is provided mainly by the muscles spindles, while information on the force exerted by the muscle, which pertains to kinetics, is provided by the Golgi tendon organs — so-called kinesthetic feedback [127].

It has long been known that humans cannot function properly without feedback.

In *Cybernetics*, Wiener [161, p.95] gives an account of a patient suffering from *tabes dorsalis*: his spinal cord has been damaged due to syphilis. As a result, his kinesthetic feedback is very limited and can marginally be compensated for through visual feedback [82] with dramatic consequences: he walks with an uncertain gait and eyes downcast, and he cannot even stand up when blindfolded.

Precise movements rely heavily on visual [82, 12] and kinesthetic [144] feedback mechanisms, especially close to the target [136]. Various experiments on occlusion and removal of light [32, 168] or removal of cursor [16] show an effect of visual feedback on virtually all kinematic properties, including accuracy and movement time. Visual feedback can have effect on movements as short as 100 ms [33, 13, 168]. Even in the absence of visual feedback, Fitts' law is usually verified [155] and movement is corrected [16] through kinesthetic feedback. Experiments on deafferented monkeys have underlined the role of kinesthetic feedback in the adaptation of learned movement [118]. Afferent signals are the signals coming from outside stimuli entering the brain. Deafferented monkeys are subjected to a surgery (e.g. bilateral dorsal rhizotomy) that selectively removes the nerve roots in the spinal cord that provide afferent signals (kinesthetic feedback) to the brain. In a condition with no limb visibility, several monkeys were trained to orient their arm towards a light. Deafferented monkeys couldn't position their arm correctly anymore when the initial position was changed from the training condition, whereas normal monkeys could quickly adapt.

Determining the minimum time needed for the Central Nervous System (CNS) to make use of feedback information has received significant attention [5, 83, 12, 13, 168, 33]. Woodworth [167] gave a first estimate of 450 ms. To account for movements shorted than about 500 ms unlikely to benefit from feedback strategies, early theories of movement control contained the concept of feedfoward control — so-called initial impulse, or ballistic phase of movement [24]. In that scenario, a motor program carefully selects a motor response [129], usually by minimizing some cost function [147]. For example, the minimum jerk principle asserts that smoothness of the movement, as measured by the mean square value of the rate of change of acceleration (jerk) is maximized by the CNS [43]. Other cost functions, such as impulse-cost, energy-cost, peak-acceleration cost [110] and duration [102, 103] are also considered.

Support for feedforward control also emerges from numerous empirical studies. Monkeys with lesions to the premotor cortex are not able to redirect their movement around a transparent object blocking the way to a reward, despite visibility of that object [164], showcasing the importance of trajectory planification before the onset of movements. D is also shown to affect the early parts of movement [78, 102]. Two consecutive movements are not independent: two consecutive undershoots occur only rarely [33], while two consecutive overshoots are even more rare [33, 117]. Certain features of the final kinematics of the movement are evident in the early kinematics [101].

Reconciling these apparent contradictory observations is easy. Indications of motor planning do not imply that the whole trajectory is initially planned; similarly evidence of feedback does not imply absence of prior planification [53]. Instead the abundant evidence for feedback and feedforward control suggests both processes occur, and

the control structure of the human can appear to be open-loop, close-sloop or a combination of both [100]. There is considerable evidence that subjects produce precise corrections based mostly on visual feedback, whereas kinesthetic feedback is likely used mostly in the early stages of movement [167, 34, 33, 102]. It is also suggested that high precision is obtained by comparing a visual signal with kinesthetic reafferent signals [120].

One could expect that with experience, the quality of feedforward control improves up to a point were it becomes sufficient to guide movements, relegating feedback information to the realm of learning. Nothing is less true, as highly trained are equally [32, 168] or even more affected [121] by a removal of feedback information than poorly trained participants. Expert participants actually learn to use feedback differently [20] and more efficiently [167, 34]

To achieve skilled motor performance, the CNS must thus exercise both closed loop control, evaluated from feedback information, and open-loop prediction [82, 101]. Successful models are expected to account for both open loop and closed loop behaviors [53, 24, 149, 34, 33].

2.1.3 Intermittent and Continuous Control

The kinematics of aiming display irregularities even when producing stereotyped movements. Fig. 2.3 displays speed profiles (speed as a function of time for individual movements) for three participants of the G-dataset performing movements towards a line under 5 different speed-accuracy conditions (from extreme speed emphasis to extreme accuracy emphasis). This paradigm differs from Fitts' Fig. 1.1 in two main respects: the target is a line, father than an interval, and the movements are made one by one (discrete protocol), rather than in a continuous alternation (discrete protocol). While some profiles seem perfectly smooth, such as the profile at the bottom right, other profiles display one or several "bumps" (e.g. third column, third and fourth row), or irregularities such as the profile in the first column and first line.

Applying a non-linear transformation to the velocity profiles allows to even better visualize these "bumps". Fig. 2.3 is replotted with the following transformation $f(x) \rightarrow \log_{10}(1 + |x|)$, which reduces the gain of high amplitude signals i.e., squeezes the top of the velocity profile, thereby making the smaller peaks more visible (in Fig. 2.4).

Velocity profiles such as those represented in Figs. 2.3 and 2.4 are produced by intermittent control [62, 22, 102, 83]; each "bump" is interpreted as a discrete correction originating from the CNS[1]. The amplitude of this correction is much smaller than the first spike in velocity. The first spike is called the primary submovement, while the other bumps are called secondary submovements — or simply submovements. Submovements have been extensively studied [78, 89], and several models explicitly use the concept [22, 102], and yet they remain surprisingly elusive, and even controversial: for example, there is no submovement visible in the bottom right panel of Fig. 2.4, even

[1]Higher order derivatives, such as acceleration or jerk may be considered in the literature. The arguments that are made for the velocity profiles transfer to these signals as well.

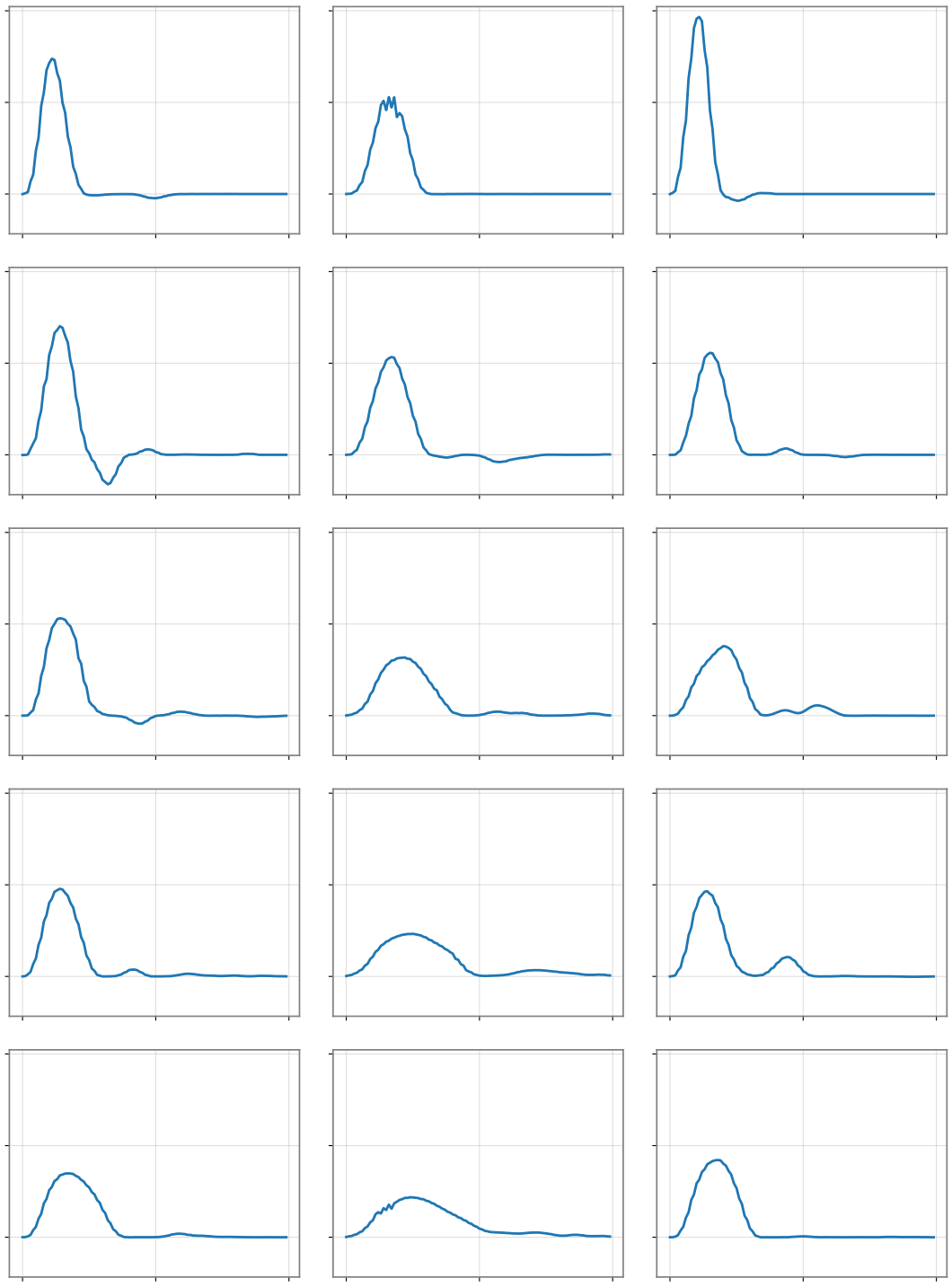

Figure 2.3: Velocity profiles (speed as a function of time) for voluntary movements from participants aiming towards a line. Each row represents a speed-accuracy condition (first row – speed emphasis, last row –accuracy emphasis). Each column is produced by a different participant. All plots are displayed with the same horizontal and vertical axes; units are arbitrary. Note that each panel corresponds to one individual movement.

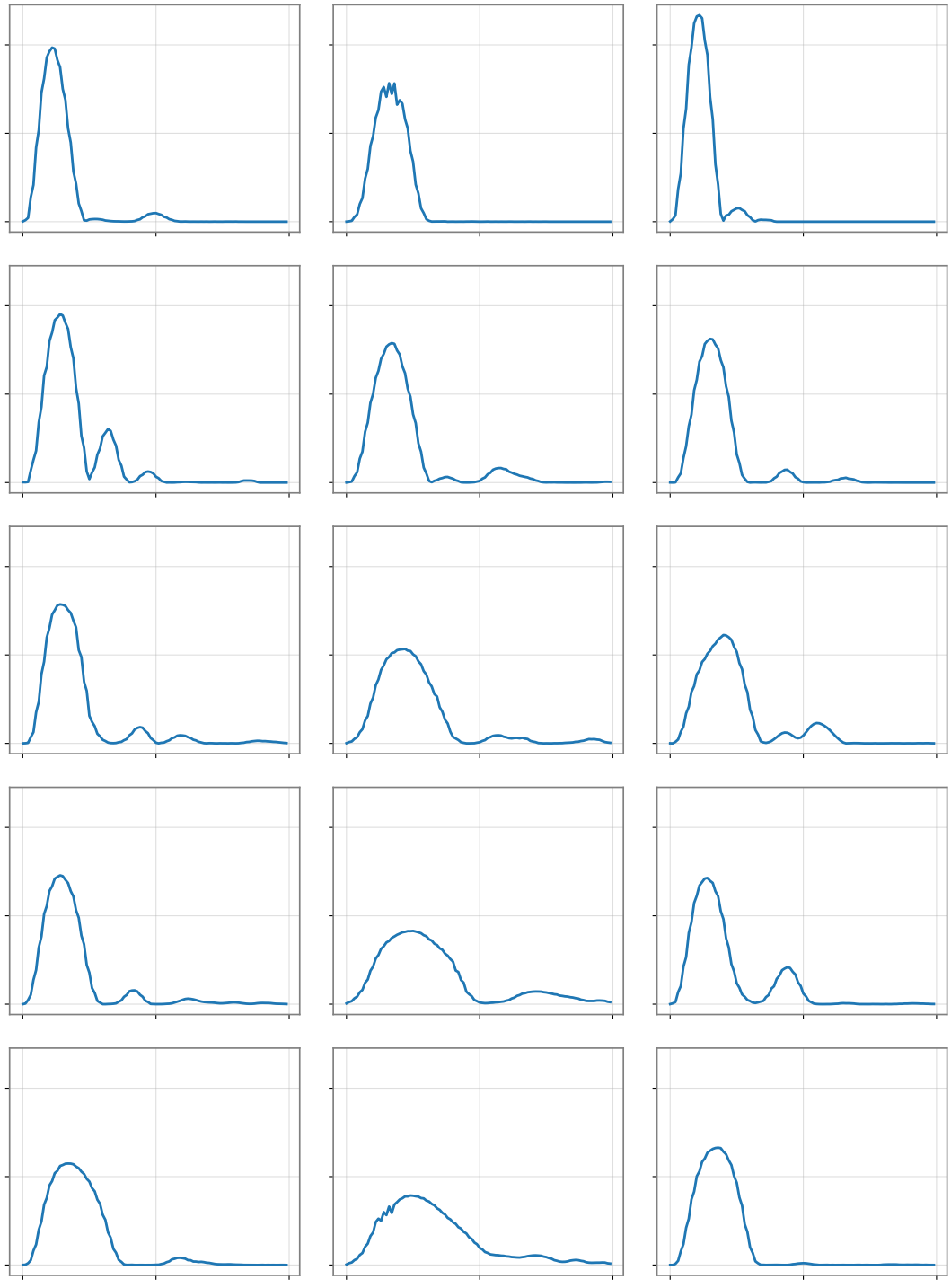

Figure 2.4: Fig. 2.3 replotted by applying the non-linear transformation $f(x) \rightarrow \log_{10}(1+|x|)$ to the velocity profile.

though this movement was produced under extreme accuracy emphasis, precisely the condition where submovements should be most present according to submovement models [22, 102]. Capturing submovements reliably is difficult for various reasons:

1. Secondary submovement amplitudes are usually much smaller than that of the primary submovement,

2. A severe low-pass filtering effect can occur due to the inertia of the limbs involved in the movement [89]. For example, some studies that are concerned with submovements explicitly target wrist movements, whose rotation involve low inertia [102, 114], but these movements are not always relevant. Movements such as those studied during pointing with a mouse involve fingers and wrist movements, while actions where target are selected by tapping on a tablet may additionally engage the upper and lower arms, and the shoulder and portions of the upper back.

3. The computation of the velocity profiles is the result of estimation, interpolating and filtering. Each of these processes may diminish the visibility of submovements. For example in the top center panel of Fig. 2.3, one can see some tremor, amplified by the differentiation process necessary to obtain velocity from position. Filtering out this tremor may unfortunately also lead to filtering of secondary submovements.

4. Segmenting submovements is arbitrary, so that two algorithms may parse a movement quite differently even with an ideal signal.[2] To our knowledge, there is no consensual or widely used parsing algorithm. Fig. 2.5 displays position and speed profiles of a single movement. The movement stops at about 27.4 s, where position is stationary (Fig. 2.5-a). However, the zoomed-in version (Fig. 2.5-b) shows it is hard to determine an exact stop time. Considering speed rather than position is no more helpful (Figs. 2.5-c and 2.5-d), as the speed never vanishes. Movements actually never stop; once a participant reaches its goal, he should still produce some effort to actually stay in place, hence some residual tremor. The transition between reaching the goal and maintaining position is usually a smooth one and occurs over several hundreds of milliseconds, a relatively long span of time compared to the total movement time. Thus, two algorithms determining the end of a movement may produce substantially different results.

Sometimes, secondary submovements are simply absent; control and feedback effects are much more continuous then a discrete sampling would imply [100] and can even appear completely continuous [115, 25, 128]. In that case, no visible changes occur in the kinematic profiles, yet the overall performance is improved due to the presence of feedback. This seems to be the case in the bottom right panel of Fig. 2.4, where velocity remains close to null after the initial spike of velocity. Numerous movement trajectories have been reported continuous, while still conforming to Fitts' law [89, 78].

[2] An extreme case is Schmidt et al.'s [129] segmentation, which leaves no time for corrections, see 2.2.3.

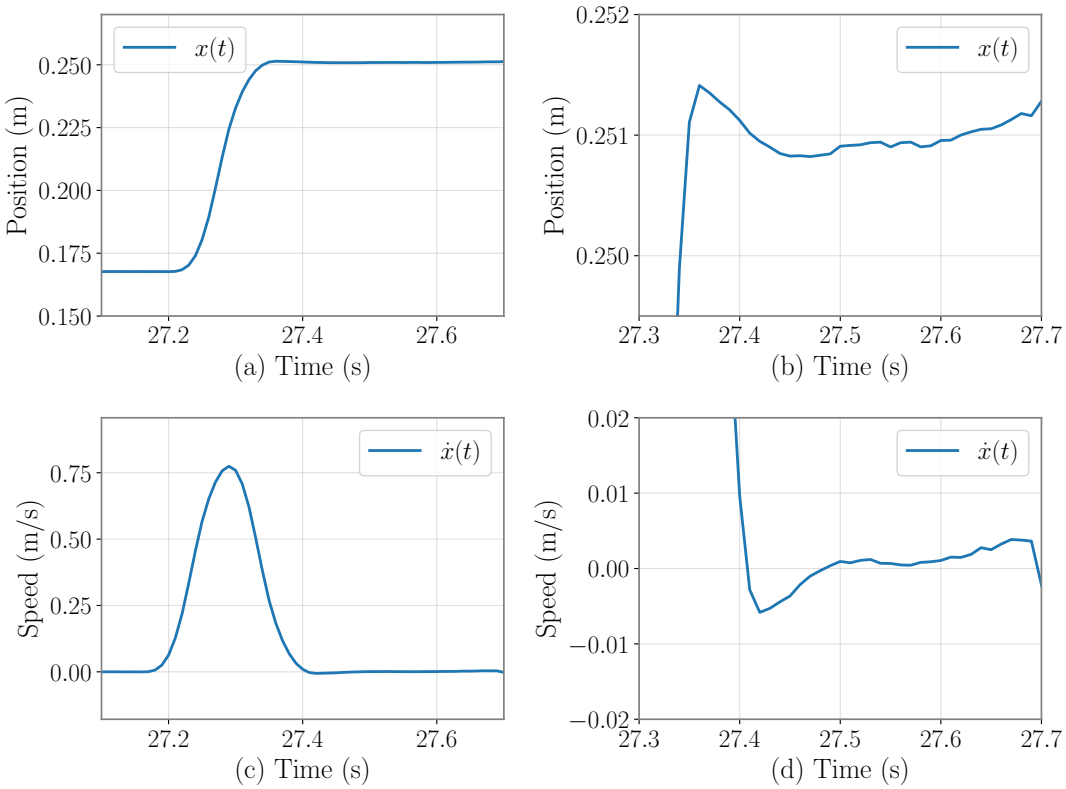

Figure 2.5: Position $x(t)$ (top) and speed $\dot{x}(t)$ (bottom) plotted versus time for a single movement of the G-dataset (see Chapter 7). On the left the full kinematic is represented, while the right displays a zoomed-in view on the end of movement.

How much this absence of visible submovements actually implies an *actual* absence of control impulses is hard to tell: Further zooming-in on Fig. 2.4 reveals that even the bottom right panel displays a small submovement, see Fig. 2.6.

Some researchers argue that compensatory adjustments cannot be identified by discontinuities in the movement profile [149], or are often overlooked [26]. Further technical difficulties arise as the number of zero-line crossings in acceleration profiles are greater when double differentation is used rather than data measured by an accelerometer [44].

2.2 Theoretical Models for Voluntary Movement

The previous section reveals that human movement is incredibly disparate; the difficulty of providing a simple model for a system capable of so much diversity is thus pregnant. We review some models for voluntary movements which share similarities with

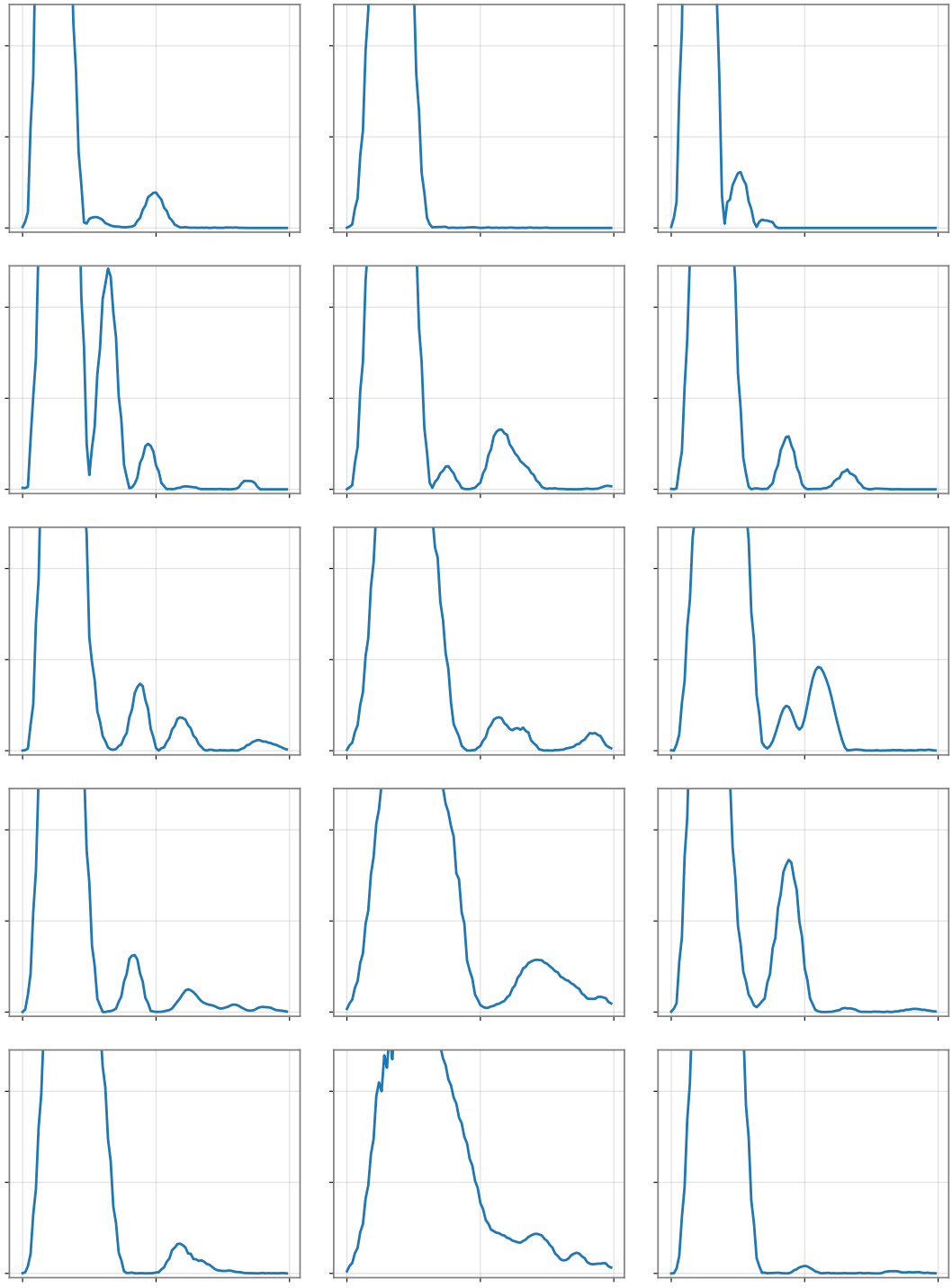

Figure 2.6: Same data as in Fig. 2.4, but with the vertical axis zoomed in by a factor 4.

our work, trying to understand where and how they succeed and/or fail. The interested reader is referred to [103, 127, 33, 117, 109] for more details and other models.

2.2.1 Early Descriptions: A Two-Component Movement

The earliest descriptions of aimed movements go back to the nineteenth century [71, p. 154], [167]. Using crafty empirical methods Woodworth [167] managed to analyze more than 125.000 movements — impressive even by today's standards.[3]

Woodworth [167] observed the speed-accuracy tradeoff when participants were allowed to see: when their speed increased, their accuracy diminished. He also found that under constant speed, increasing the distance to be covered would lead to a decrease in accuracy, while shorter time intervals between movements favored accuracy.

Based on these findings, Woodworth [167] hypothesized that two serial components constituted any aimed movement: an *initial adjustment*, whose main purpose is to cover distance, followed by *current control*, which ensures that accuracy constraints are met. He demonstrated the existence of re-adjustments following the initial adjustments and extensively discussed the role and impact of visual and kinesthetic feedback on movement, as well as fatigue and learning.

2.2.2 Crossman and Goodeve's Deterministic Iterative Corrections (DIC) Model

Crossman and Goodeve [22] presented a feedback interpretation of Fitts' law, as they doubted the existence of the concept of noise that Fitts had leveraged (see 1.2.2).

Crossman and Goodeve [22] proposed a scheme of intermittent proportional correction of position: If x_k is the distance from the center of the target at iteration k, then

$$x_{k+1} = x_k - Mx_k, \quad M \in]0;1[. \tag{2.1}$$

The scheme is called proportional, as the magnitude of the correction Mx_k at each time is proportional to the actual error x_k. Setting the initial distance to D ($x_0 = \mathrm{D}$), the distance at iteration k is the geometric closed form expression

$$x_k = \mathrm{D}\big(1 - M\big)^k. \tag{2.2}$$

Assuming a constant time interval δt between each correction, the movement time MT is computed as $\mathrm{MT} = k\ \delta t$. Assuming that the movement stops as soon as possible i.e. $x(\text{end}) = \mathrm{W}/2$, and from (2.2), MT is obtained as

$$\mathrm{MT} = -\delta t \log_{1-M} \frac{2\mathrm{D}}{\mathrm{W}} = -\frac{\delta t \log(2)}{\log(1-M)} \log_2 \frac{2\mathrm{D}}{\mathrm{W}}, \tag{2.3}$$

[3] His analyses were usually accompanied by a measure of statistical significance, the *probable difference*, long before Fischer popularized the p-values. Woodworth also observed and commented upon the confound between planification time and movement time in a serial tapping task. These are some of the elements that make Woodworth's work truly remarkable. Unfortunately, Woodworth's work is undervalued and he is himself quite unknown these days.

Formula (2.3) is equivalent to (1.17), but without intercept (a), and with $b = -\frac{\delta t \log(2)}{\log(1-M)}$. This is always a positive quantity as $\log(1-M)$ is negative for $M \in]0;1[$.

The major limitation of this model is that it is deterministic, which is incompatible with the variability typically found in human motor movement (1.2.2), and the predictions of the DIC model are falsified by many empirical studies e.g. [102]. Nonetheless, it was the first motor theory to hypothesize a control of position through intermittent feedback, which explains why it remained popular for quite some time. Moreover, the idea that the difference between target and limb position plays an important role in the control of movement and the generation of control impulses has received significant empirical support [33]. We will make use of this mechanism later on.

2.2.3 Schmidt's Law

Using a time-constrained rather than a width-constrained task and studying the variability of the human response, Schmidt et al. [129] observed the so-called linear speed-accuracy tradeoff for rapid movements:

$$\mathrm{MT} = a' + b'\mathrm{D}/\sigma, \tag{2.4}$$

where MT and D are defined as in Eq. (1), and σ is the standard deviation of endpoints.

The discrepancy between Schmidt's and Fitts' law is most likely due to the method used by Schmidt and colleagues to determine the end of the movement which leaves no chance for late discrete control [33]. In fact, Schmidt et al. described considerable residual deceleration after what they considered to be the end of a movement. In practice, Schmidt's law is often seen as a formula for ballistic open loop movements of short duration, whereas Fitts' law is considered to be the operational formula for aimed closed loop movements of longer duration.

2.2.4 Meyer et al.'s Stochastic Optimized Submovements (SOS) Model

Woodworth's idea of a two-component movement, Schmidt's linear speed-accuracy tradeoff for ballistic movements and Crossman and Goodeve's DIC model are combined in the SOS model. Each movement is assumed to begin with a primary submovement which is ballistic, whose role is to cover most of the distance separating the initial point from the target. It is thus very close to Woodworth's description of the distance-covering component. In SOS, the primary submovement of duration t_1 follows Schmidt et al.'s linear tradeoff (2.4) without intercept. Once this primary submovement is finished, either the endpoint is inside the target, in which case the movement stops, or the endpoint is outside the target and a corrective submovement is needed. This corrective submovement has exactly the same behavior as the primary one, except that its variability is naturally lower because the distance to be covered is usually much shorter; the time needed to perform this secondary submovement is t_2. By minimizing the sum $t_1 + t_2$, the SOS model predicts that movement time is given by

$$\mathrm{MT} = a + b\,(\mathrm{D}/\mathrm{W})^{1/2}. \tag{2.5}$$

This model was later extended to the case of $n \geq 2$ submovements [103]. In the limiting case, when n grows large, it was shown that $(\mathrm{D}/\mathrm{W})^{1/2}$ should be replaced by the logarithmic law $\log_2(\mathrm{D}/\mathrm{W})$ [124]. Thus, in the limit where the number of submovements is allowed to grow large enough, the SOS model predicts MT in agreement with Fitts' formula (1.17). However, one can question the relevance of such an asymptotic result for human movement as Meyer et al. [102] rarely report more than 4 submovements.

Fitts' law is thus expressed in SOS as the result of an optimal allocation of time between a ballistic primary submovement and a succession of corrective submovements. In essence, the stochastic optimized submovement model is a stochastic version of Crossman and Goodeve's model, where each individual submovement follows Schmidt's law. As it effectively combines elements of several successful models into a comprehensive description, the SOS is one of the leading explanatory models for Fitts' law [127, 33].

However, the SOS model suffers from several well-documented deficiencies [102, 34], including some theoretical predictions on submovement that do not match empirical findings [34]. For example, D and W being fixed, there is considerable variation in the duration of the first submovement [89, 78, 117], and the number of submovement required on average for a given task is poorly predicted. More importantly, the model is unable to account for continuous feedback; as a result its empirical validation relies on identifying submovements, with all the caveats described in 2.1.3.

2.2.5 Bullock & Grossberg's Vector-Integration-To-Endpoint (VITE) Model

A neural network model called VITE was proposed [8] on the basis that the complicated synergies between muscle groups exclude a pre-planning (ballistic) strategy. The model builds on three signals: target Position Command (TPC) which is used to locate the target, present Position Command (PPC) which is used to locate the current position of the arm and a GO command, used to modulate the overall speed of the movement. At any time instant, the difference between TPC and PPC, corresponding to the remaining distance to be covered is evaluated into a difference vector (DV) which is then multiplied by the GO signal and integrated to form the new PPC. DV is a vector, as each of its component corresponds to the difference between TPC and PPC for a specific synergetic muscle group, and is the signal (when multiplied by GO) that drives the agonist and antagonist muscle groups. The GO signal on the other-hand is non-specific and can assume several shapes, including step functions and sigmoid functions.

The systems equations reveal that VITE is a non-linear time-continuous second-order system, with no simple close form solution. However as the equations describing the model are deterministic differential equations, the model fails to account for variability[4].

Furthermore, the model's behavior is largely dependent on the shape of the GO

[4]It can always be argued that the parameters fitted to the model can vary stochastically, making any deterministic model somewhat stochastic. This is not a convincing method to explain the consistency and flexibility of human movement [99]

signal, which is determined *a posteriori* to fit existing trajectories, but for which we have no *a priori* knowledge.

2.2.6 Plamondon & Alimi's $\Delta\Lambda$ model

Plamondon & Alimi's [117] model is based on two observations: that asymetric velocity profiles are an important invariant of human aimed movement and that a movement is produced by the synergy of agonist and antagonist muscle groups. In $\Delta\Lambda$, each system (agonist and antagonist muscle groups) is modeled by a large number of sequential linear time-invariant (LTI) processes. If each process j induces a time delay proportional to the cumulative delay of the previous $j-1$ processes T_{j-1}, the total time delay at process j can be given iteratively:

$$T_j = T_{j-1} + \varepsilon_j T_{j-1} = (1+\varepsilon_j)T_{j-1}. \tag{2.6}$$

Under weak constraints, Plamondon et al. then show that the impulse response of this sequence of LTI processes converges towards a lognormal curve. The weighted difference of two of these impulse responses (each impulse response representing the response of a group of either agonist or antagonist muscles), gives the velocity profile, hence the term delta-lognormal (delta for difference). The predicted velocity profiles are remarkably close to real-life measured kinematics, although admittedly the model was conceived with precisely this goal and the number of free parameters (seven) is quite large.

Although based on the idea of random time lags, the model produces deterministic velocity profiles, and unfortunately fails to incorporate feedback, as all the movements are the result of pre-planned impulse responses. An interesting feature of this model is that in some cases, a single impulse can lead to a velocity curve with two spikes, which further underlines the difficulty of submovement parsing.

Its great fitting power for velocity profiles have made it a useful tool for modeling the generation of complex trajectories such as those involved in writing or making signatures, but fails to capture most of the required aspects of aimed movement.

2.2.7 Elliott and Colleague's Two-Component Model

Building on extensive experimental accounts, Elliott and colleagues [34, 33] have proposed a description of aimed movement inspired by the two component model of Woodworth [167]:

1. a first *planned* component, that gets the limb close to the target area. This planification of movement is based on internal models and representations but, contrary to e.g., Schmidt's [129] description, it is not uniquely determined by forward control. Instead, it is associated to a velocity regulation through mostly kinesthetic/proprioceptive feedback.

2. When time permits, a second corrective portion is engaged to reduce any spatial discrepancy between limb and target. This process, highly dependent on foveal (central) vision, involves computing the difference between limb and target position and issuing discrete corrections. With practice, these corrections may appear smooth and continuous.

At present, there is unfortunately no quantitative model for this two-component model.

2.2.8 Discussion

None of the models above capture all the characteristics of human aimed movement described in 2.1. Deterministic models such as DIC and VITE fail to account for the variability of human movements and $\Delta\Lambda$ does not include ways to integrate feedback. VITE does not account for intermittent control nor feedforward mechanisms, while the celebrated SOS model fails to consider continuous corrections, yet characteristic of skilled operators [167, 115]. The most precise description is probably due to Elliott and colleagues [34, 33, 35], inspired from Woodworth's [167] two component model, but it lacks a quantitative model that provides operational results. A control theoretic model by Todorov et al. [147] represents the human as a Kalman filter driven by an optimal controller, and accounts for the diversity discussed in 2.1. However, it seems poorly adapted to produce results in HCI [109], the biggest drawback being that an input of the model is MT i.e. a quantity that we usually need at the output of a model.

Chapter 3

Fitts' law: Methods and Controversies

3.1 Regressing Movement Time

3.1.1 Time Metrics

Fitts' formal hypothesis in 1954 [38] was that: *" If the amplitude and tolerance limits of a task are controlled by E [the experimenter], and S [the participant] is instructed to work at his maximum rate, then the average time per response will be directly proportional to the minimum amount of information per response demanded by the particular conditions of amplitude and tolerance "* [38, p. 2]. The formula for MT given by Fitts (1.15) is expressed as the ratio between a nominal quantity (ID) and C, the Shannon capacity, representing the participant's *maximum rate*. If C is a maximum, the movement times should correspond to minimum movement times. Yet, we should, according to Fitts, consider the average of those minima over all time measures. Fitts' MT metric is thus an average of the minimum movement times. This raises three issues:

- Evidently, the oxymoron of an average minimum described by Fitts is confusing. Several authors use, instead, different wordings, which suggest other interpretations of MT. A few examples are given by Soukoreff et al. [142] who consider "movement time performance for rapid aimed movements", Hoffman [75] "movement time", and Drewes [27] "mean time".

- Fitts needs the participant (S) to work at his or her maximum rate, so that the resulting movement times reflect S's full commitment to the pointing task. Yet, subjects are rarely fully committed to repetitive tasks such as those composing Fitts' pointing experiments. The use of the average of MT is thus based on a false premise, resulting from a shortcoming of the experimental design. As emphasized by Guiard et al. [66, 67, 57] It is unreasonable to take it for granted that the participants routinely perform to their fullest capabilities during the course of an experiment, even if they were instructed to do so.

- Fitts' average-minimum metric cannot be defined in a field study, because the participant are never instructed to operate at their maximum rate.

3.1.2 Regression

To convert empirical data to a meaningful model, researchers don't usually regress the pointing data directly. Instead, they compute block averages first (i.e. they aggregate all the data from one block, and compute the average movement time for each block); only then is the regression computed.

Fig. 3.1 displays a sample dataset, taken from [81], produced in a Fitts task. The top panel shows the entire dataset, while the bottom panel shows averaged data, in orange diamonds. Usually, papers only display the orange diamonds, along with the linear regression line. Each dot represents one trial i.e. movement time for one movement acquired with the corresponding level of ID. The set of dots constitute a single run (D3P09) of 10 conditions.

We performed a simple Least Squared Estimate (LSE) procedure on the dataset with Fitts' model (1.17) on both panels. For the top panel, the r^2 found is well below what is typically displayed by researchers, while for the bottom panel, r^2 is above 0.9, in line with typical values [142].

Although the model parameters a and b estimated by LSE barely change, the r^2 increases dramatically. Because the difference between the two datasets has only been the averaging procedure for data of the bottom panel, the increase in r^2 is due to the reduction of within-subject variability by averaging.

The uncertainty concerning the exact nature of the MT, as discussed just above leads to two models for movement time:

1. Fitts' law, or the "post-average law" as typically computed through LSE on averaged data but expressed as

$$\overline{\mathrm{MT}} = \overline{a} + \overline{b}\,\mathrm{ID}, \tag{3.1}$$

   where the overline refers to average, and

2. The "pre-average law", as typically expressed

$$\mathrm{MT} = a + b\,\mathrm{ID}, \tag{3.2}$$

   but computed through LSE on the whole dataset, prior to averaging.

If we are interested in elementary trial time, then considering averages (i.e. fitting (3.1) but wanting to describe (3.2)) is generally speaking an instance of aggregation bias [68] i.e. thinking that relationships observed for groups necessarily holds for individuals — the so-called ecological fallacy [45].

A relevant example to aimed movements is given by Holly [117, p. 313–314], who explains that the average data computed from data that follows the $\Delta\Lambda$ model (see 2.2.6) will not generally follow the $\Delta\Lambda$ model.

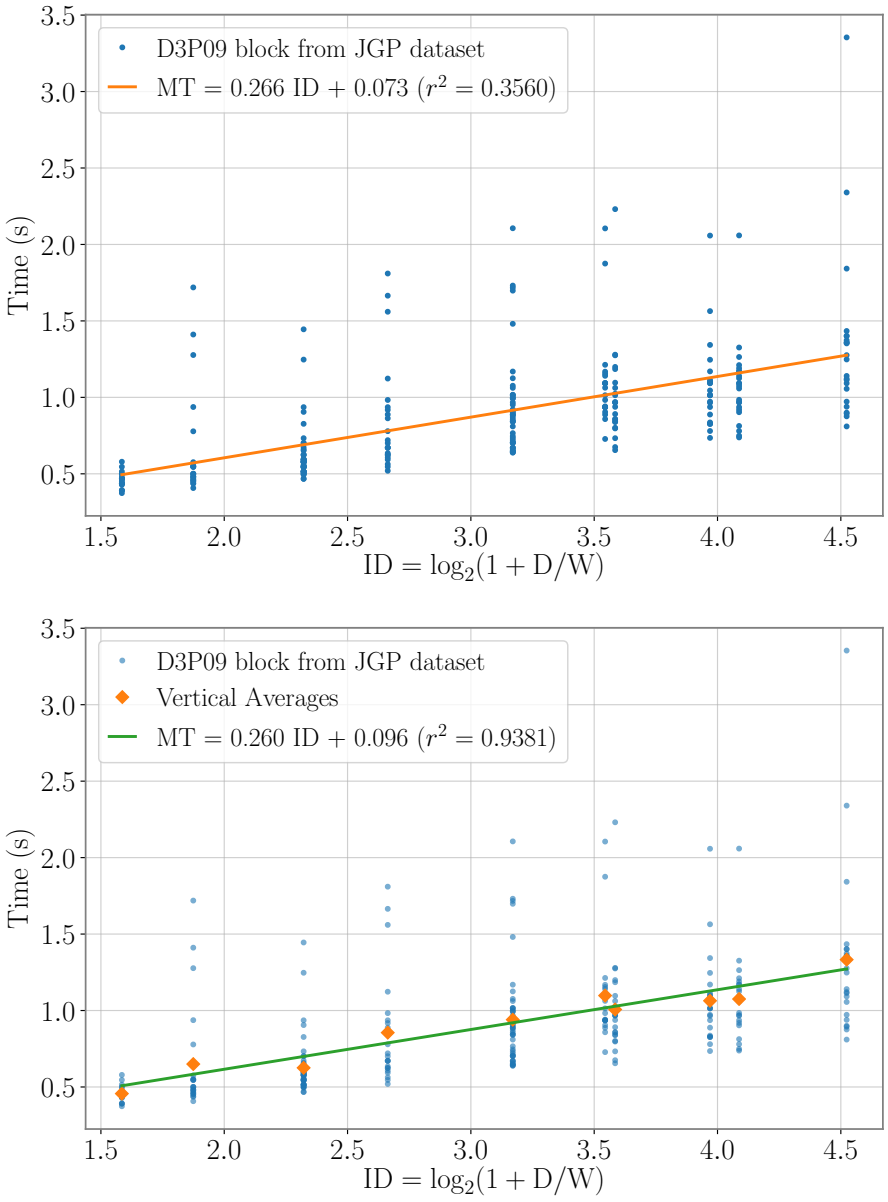

Figure 3.1: Pointing data from the JGP dataset (see Chapter 7). An identical block is considered in both panels. The top panel shows a linear model fitted by LSE to the whole data. The bottom panel shows a linear model fitted by LSE to the mean of the data per condition.

Aggregating and averaging data can also result in a confirmatory bias [58], where Fitts' law becomes indistinguishable from other competing laws.

In spite of all these caveats, the post-average law successfully predicts $\overline{\mathrm{MT}}$. However, it is clear from Fig. 3.1 that (3.2) is a really poor model for movement time. The concept of front of performance, discussed in Chapter 5 will be useful to capture minimum movement time and will be leveraged further in Chapter 8.3.2 to provide a model for movement time for the whole dataset.

3.2 Formulation

The theoretical basis of Fitts' law has been criticized from the onset, and many models can produce a law similar to Fitts'. As a result, over a dozen formulations have emerged [117]. We give a sample of some of the most known:

Fitts [38]	$\mathrm{MT} = a + b\ \log_2\big(2\mathrm{D}/\mathrm{W}\big)$	(3.3)
Crossman [21]	$\mathrm{MT} = a + b\ \log_2\big(\mathrm{D}/\mathrm{W}\big)$	(3.4)
Welford [157]	$\mathrm{MT} = b\ \log_2\big(1/2 + \mathrm{D}/\mathrm{W}\big)$	(3.5)
Jagacinski et al. [21]	$\mathrm{MT} = a + b\,\mathrm{D} + c\ \big(1/\mathrm{W} - 1\big)$	(3.6)
MacKenzie [93]	$\mathrm{MT} = a + b\ \log_2\big(1 + \mathrm{D}/\mathrm{W}\big)$	(3.7)
Gan and Hoffmann [48]	$\mathrm{MT} = a + b\ \sqrt{\mathrm{D}}$	(3.8)
Kvålseth [86]	$\mathrm{MT} = a\ \big(\mathrm{D}/\mathrm{W}\big)^b$	(3.9)
Welford et al. [159]	$\mathrm{MT} = a + b\ \log_2\mathrm{D} + c\ \log_2\mathrm{W}$	(3.10)

In HCI the so-called Shannon formulation (3.7) is overwhelmingly used, contrary to e.g. experimental psychology [107, 117] where Fitts' original formulation (3.3) is commonly considered.

> *The superiority of the Shannon [MacKenzie] formulation over the two other popular formulations, those of Fitts* (3.3) *and Welford* (3.5)*, is well documented [93, 94, 95]. The Shannon formulation is preferred because: (i) it provides better fit with observations (a higher correlation-coefficient is typically achieved), (ii) it exactly mimics the information theorem that Fitts' law is based on, and (iii) with this formulation negative* ID *values are not possible.* [142]

As we will show, these 3 items are all disputable, starting with (ii). As detailed in 1.3.1, Fitts' formulation and by extension Mackenzie's formulation are just vague analogies.[1] The two other items are discussed below; but beforehand we give a simple remark on the equivalence between some indexes.

[1] Mackenzie's argument [93] is imprecise anyways when he claims Fitts makes the analogy between D and P, as in reality Fitts [38, 40] identified D to $P + N$.

3.2.1 A Remark on the Equivalence Between Indexes

Proposition 3.2.1. *Suppose that we have two models for* ID *such that* $\mathrm{ID}_1 = \alpha + \beta \cdot \mathrm{ID}_2$. *Then both will be equivalent for Fitts' law, in the sense that the two models will have the exact same fitting power.*

The proof is trivial as $\mathrm{MT} = a_1 + b_1\ \mathrm{ID}_1 = a_1 + \alpha b_1 + \beta b_1\ \mathrm{ID}_2 = a_2 + b_2\ \mathrm{ID}_2$. Because the $a_\cdot$ and $b_\cdot$ constants have to be estimated from empirical data points, both indexes are equivalent. The values of a and b are also easily interpretable as one can go from the parameters of one description to the other by simple calculations.

For example, Fitts' index [38] $\mathrm{ID} = \log_2(2\mathrm{D}/\mathrm{W}) = 1 + \log_2(\mathrm{D}/\mathrm{W})$ is equivalent to Crossman's index [21] $\mathrm{ID} = \log_2(\mathrm{D}/\mathrm{W})$. Also, the "mixed" Fitts-MacKenzie's expression $\mathrm{ID} = \log_2(1 + 2\mathrm{D}/\mathrm{W})$ is equivalent to Welford's index since $\log_2(1 + 2\mathrm{D}/\mathrm{W}) = 1 + \log_2(1/2 + \mathrm{D}/\mathrm{W})$.

As another illustration, consider the novel formulation for ID proposed by Soukoreff et al. [143]:

$$I_{\mathrm{D}}\ \text{entropy} = m + \log_2(U) - \frac{1}{2}\log_2\Big(\pi e \frac{\mathrm{W}^2}{8}\Big) + 1, \tag{3.11}$$

where U is the "size of the movement universe", i.e. the largest extent considered for movements. Grouping the logarithms together, we obtain

$$I_{\mathrm{D}}\ \text{entropy} = m + 1 + \log_2\Big(\frac{2U}{\mathrm{W}}\sqrt{\frac{2}{\pi e}}\Big) = m + 1 + \log_2\Big(2\sqrt{\frac{2}{\pi e}}\Big) + \log_2\Big(\frac{U}{\mathrm{W}}\Big). \tag{3.12}$$

Now considering the largest extent to be either D, $\mathrm{D} + \frac{\mathrm{W}}{2}$, or $\mathrm{D} + \mathrm{W}$, one recovers the indices of difficulty of Fitts, Welford, and Mackenzie, respectively.

3.2.2 Fit of the Mackenzie Formulation

As described by Roberts & Pashler [126], a theory cannot be justified by any single empirical fit. However, many quantitative theories commonly use a good fitting power to account for their models, and the last quote reveals that Fitts' law is unfortunately not an exception.

Furthermore, reasonable fits are not particularly surprising as Fitts' law is replete with statistical methods for reducing variability [117], such as using LSE after averaging. As a means of discriminating between formulations fits are thus even less significant, because most of the variability of the data is concealed. To see this, consider the dataset used in Fig. 3.1, and compare the fits produced by the formulations of Fitts (3.3), Welford (3.5), Mackenzie (3.7) and Gan and Hoffmann (3.8). We have computed the fits pre and post averaging of the data, see Table 3.1. Gan and Hoffmann's formulation (3.8) produces the best fit on the post averaged data, yet produces the worst fit on the pre averaged data.

The three other formulations are very close in their goodness of fits, and we should rightfully conclude that all alternative theories do roughly equally well in predicting the

actual movement time needed for different targets [150]. This, however, is just one example. To generalize, we would have to iterate this procedure multiple times (see [75] and [96]). An elegant alternative exists, in the form of Akaike's Information Criterion (AIC) [2, 9]. AIC is a metric that aims to quantify how good a model is with respect to the unknown, ground truth, based on the information-theoretic Kullback-Leibler divergence, which essentially quantifies the "distance"[2] between two probability distributions, i.e. two models. It is commonly used to select a model that fits data adequately, but by considering the risk of overfitting by penalizing a model that has too many free parameters. In the case of a univariate linear model with Gaussian homoskedastic noise (i.e. the most usual model), the AIC criterion is simply written as [9]

$$\mathrm{AIC} = n\log(\widehat{\sigma}^2) + 2K = n\log\frac{\mathrm{RSS}}{n} + 2K, \tag{3.13}$$

where n is the number of sample points, $\widehat{\sigma}^2$ the Maximum Likelihood Estimate (MLE) of the variance of the residuals[3], RSS the Residual Sum of Squares and K the number of parameters that are fitted. A better AIC indicates a better fit with the ground truth, although the value can not be interpreted as is; only differences in AIC are meaningful. Note that the smaller the RSS and the smaller the number of parameters fitted, the better; AIC thus provides an operational solution to the well known observation that adding parameters to a fit increases fitting power but decreases explanatory power [9].

To treat competing models [9], one should compute all AIC's, find the minimum AIC_{min}, compute the AIC differences

$$\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{min} \tag{3.14}$$

and from there compute the likelihood of each model

$$\mathcal{L}_i \propto \exp\left(-\frac{1}{2}\Delta_i\right). \tag{3.15}$$

The likelihood of each model should then be interpreted as odds [9], i.e. there is a $\exp\left(-\frac{1}{2}\Delta_i\right)$ in one chance that the model with Δ_i is closer to the ground truth than the one with AIC_{min}. The odds for each model are given in Table 3.1. The beneficial aspect of the AIC appears clearly, as we are now able to tell whether a difference in r^2 is meaningful or not: in this case it is clear that all models are essentially equivalent.

It is beneficial to reflect on the many models (2.2) leading to Fitts' law and its many formulations (3.3)–(3.9) to understand why so many diverging models [150] seem to converge towards it.

The basic principles of the law are rather unsurprising, and it is reasonable to expect even a toddler to understand the simple fact that picking up a large object close to him is easier than picking up a tiny object that he can just about reach. In a slightly more

[2]The Kullback-Leibler divergence lacks the symetry property to be a distance in the mathematical sense, see e.g. [18].

[3]The MLE estimate differs slightly from the MMSE estimate ($\mathrm{RSS}/(n-2)$) computed for linear regression.

Table 3.1: LSE fits for different formulations of Fitts' law

	Fitts	Welford	MacKenzie	Gan & Hoffmann
pre average r^2	0.3560	0.3562	0.3560	0.2742
post average r^2	0.9398	0.9393	0.9381	0.9791
pre average AIC odds	0.97	1	0.95	
post average AIC odds	1	0.95	0.86	

precise form, namely that MT is an increasing function of D and decreasing function of W, the law does not become much more surprising.

However, the fact that the important variable is not D, nor W, but the ratio $\mathrm{D/W}$ is in itself surprising, and this single fact constitutes in my opinion the major finding of Fitts. This translates to the fact that e.g. Fitts' formulation (3.3) is far stronger than e.g. the formulation of Welford et al. (3.10).

Once this is acknowledged, any concave function f with $f(0)$ close enough to 0 will likely fit any pointing data, when the parameter controlling the rate (i.e. the b in (3.3)–(3.9)) is left up to the experimenter to be decided. This explains why a square root (3.8), a logarithm (3.3) or a power function (3.9) all provide reasonable fits. Even a linear trend can appear to reasonably fit pointing data, when the number of data points are small and if the scales chosen on ordinates and abscissae bias the observation [71].

It is therefore safe to say that any of the mainstream formulation is appropriate from a data fitting perspective. There is not much to be gained for HCI from quarreling on which formulation is best; all formulations will likely lead to similar operational results.

What is worth exploring is whether we can find a model that explains the data pre averaging. We will see later that Chapter 8.3.2 offers an answer to this issue.

3.2.3 Small Values of ID

The third argument of Soukoreff & Mackenzie [142] to accept the Mackenzie formulation (3.7) relates to small values of ID. Indeed with Fitts' formulation (3.3), negative values can occur; according to Guiard [65] this is however not a problem since ID is not measured on a ratio scale (the value $\mathrm{ID} = 0$ is not properly defined).

For small values of ID (below $\mathrm{ID} = 2$ measured in (3.3) i.e. $\mathrm{D} = 2\mathrm{W}$), Crossman & Goodeve [22] modeled movement time as constant (190 ms), see Fig. 3.2.

It is well known that Fitts' formulation (3.3) fails at low levels of ID [38, 48, 93]. This is probably due to the fact that no feedback is required for movements of ID smaller than about 3 [48] (measured with (3.3) i.e. $\mathrm{D} = 4\mathrm{W}$). This is supported by evidence that vision is only needed beyond the level $\mathrm{ID} = 3.58$ [155] (measured with (3.3) i.e. $\mathrm{D} \simeq 6\mathrm{W}$). Hence, the MacKenzie formulation (3.7) is just a "hack" to Fitts' formulation, the goal being to extend the fitting power of the law a bit further towards low levels of ID. Although the hack is convenient, no theoretically useful conclusions can come out of it. In Chapters 6 and 9, we will document this view further.

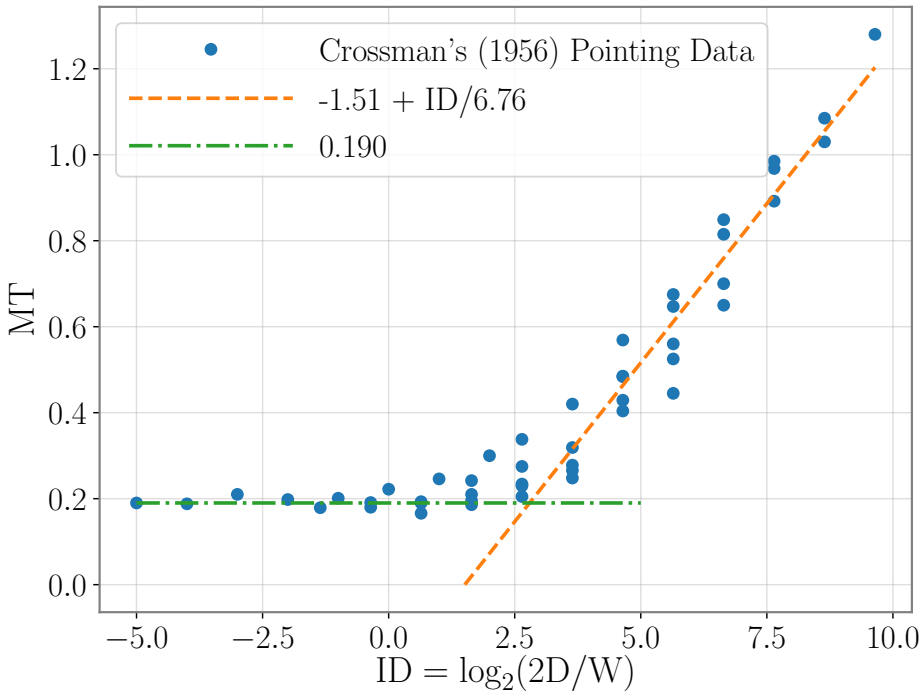

Figure 3.2: Pointing data from Experiment I in [20] with fits computed in [22]. For low levels of ID, MT is fit with the green dot-dashed line, whereas for higher levels, MT is fit with the orange dashed line.

3.3 Nominal versus Effective Width

The mathematical expressions (3.3)–(3.10) are formulated with D and W. These are fixed parameters, decided at the onset of the experiment, and constitute what are called the *nominal* formulations for ID. The experimenter relies on the participant to actually perform movements that travel a distance D and end up in the target W.

This is an expectation however that no participant can achieve in practice, except for very easy conditions [38, 20, 39, 158, 142]. Participants also have different strategies; some will be conservative and try to make as few errors as possible while others will try to reduce movement time without caring too much for precision [67]. The nominal formulation is thus hard to adapt to the execution of human movements.

Crossman [20, 21] was the first to introduce a correction for accuracy, by using a measure of the actual spread of endpoints rather than W to measure accuracy; the current practice in HCI [142] is to correct both D and W to reflect effective performance

by the user with D_e and W_e (but see also [171]):

$$\mathrm{D}_e = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{3.16}$$

$$\mathrm{W}_e = 4.133\sigma, \tag{3.17}$$

where σ is the sample standard deviation, d_i the actual endpoint distance at trial i, and n the total number of trials. The *effective* index of difficulty ID_e is then defined as

$$\mathrm{ID}_e = \log_2\big(1 + \mathrm{D}_e/\mathrm{W}_e\big) \text{ and gives} \tag{3.18}$$

$$\mathrm{MT} = a + b\ \log_2\big(1 + \mathrm{D}_e/\mathrm{W}_e\big) \tag{3.19}$$

There are important conceptual and practical differences between the two indexes [171]. In contexts where a researcher or designer is trying to make a prediction (e.g., provide a theoretical estimate of selection time in a GUI), the nominal law (1.17) is more adapted, as it is expressed from the variables that describe the problem. However, it makes sense to use a metric such as ID_e in cases where a researcher investigates the user's actual performance.

The effective index ID_e is supposedly justified by information-theoretic principles [94, 142], and defined as follows (replacing D with D_e has very little effect and we will focus solely on W_e). Let σ denote the standard deviation of the end-point distribution, and ε the error rate, i.e., the proportion of target misses:

- If σ is available:

$$W_e = 4.133\ \sigma. \tag{3.20}$$

- Otherwise:

$$W_e = \begin{cases} W \cdot \frac{2.066}{z(1-\varepsilon/2)} & \text{if } \varepsilon > 0.0049\% \\ 0.5089 \cdot W & \text{otherwise.} \end{cases} \tag{3.21}$$

According to MacKenzie [94]:

> "The entropy (H), or information, in a normal distribution is $H = \log_2\big((2\pi e)^{\frac{1}{2}}\sigma\big) = \log_2(4.133\sigma)$, where σ is the standard deviation in the unit of measurement. Splitting the constant 4.133 into a pair of z-scores for the unit-normal curve (i.e., σ = 1), we find that the area bounded by z = $\pm$2.066 represents about 96 % of the total area of the distribution. In other words, a condition that target width is analogous to the information-theoretic concept of noise is that 96 % of the hits are within the target and 4 % of the hits miss the target [...]. When an error rate other than 4% is observed, target width should be adjusted to form the effective target width in keeping with the underlying theory."

This methodology raises two issues:

1. To our knowledge Information Theory provides no justification to the relation $W_e = 4.133\ \sigma$. When Crossman [21] calculated the expression for W_e from the area under the standard normal curve, he took the 5% value as an arbitrary "permissible" error rate. MacKenzie [94] noticed that by changing the arbitrary rate from 5% to 3.88% (approximately 4%), the entropy of the rectangular distribution of width W_e would equal the entropy of the Gaussian distribution of standard deviation σ (see *Computation* 1 in 3.6), but this is no more than a coincidence—we can see for now no information-theoretic reason to equalize these two entropies. Incidentally, these entropies can both be negative, as they are differential entropies, and thus cannot be interpreted as a measure of information [18].

2. The threshold of error rate placed at 0.0049% in (3.21) is arbitrary. The one-to-one relationship between standard deviations and error rates is only true for strictly positive error rates. Indeed, when the error rate vanishes, so does the standard deviation, and so $\mathrm{ID_e}$ tends to infinity. To prevent this from happening, Soukoreff and MacKenzie [142] have recommended that below a certain error rate (0.0049%), $\mathrm{ID_e}$ should be kept constant. The justification of the threshold error rate of 0.0049% is that it "rounds to 0.00". As we will show below, the existence of such a threshold and its value of 0.0049% are in fact adverse to the theory.

The standardized index of difficulty $\mathrm{ID_e}$ is thus questionable, as it relies on two arbitrary constants and one coincidence. Even more importantly, it has never been shown to be the expression of the capacity of a human-motor channel—the expected rationale behind Fitts' law within the information-theoretic framework.

We now provide the computation that leads to W_e for completeness.

The empirical basis of ID_e is also criticized [171] (although not as much as its theoretical basis), and the effective law is not systematically used in HCI [166, 142], even when evaluating user performance (e.g. [80]). Ironically, the effective formulation (3.19) does not generally produce improved fits [94] contrary to what was intended.

However, it facilitates comparisons across experimental factors and across studies [94, 171], as a and b in (3.19) vary less across different paradigms than a and b in the nominal (1.17). This may explain why the effective ID_e is preferred over the nominal ID in the ISO standard [1] designed for the evaluation of input devices.

A final point is worth discussing. In going from the nominal formulation to the effective formulation, D and W lose the status of independent variable: as D_e and W_e depend on the performance of the user, they are dependent variables. It follows that:

- Goodness of fit cannot be evaluated solely from the r^2 value. Indeed, when fitting empirical data to (3.19), W_e and D_e have to be estimated. The free parameters are thus 1) the two line parameters a and b 2) the 12 values of W_e (one for each condition) 3) the value of D_e 4) the variance of the residuals, which makes 13 extra parameters compared to the nominal formulation. Put this way, obtaining equivalent fits at the cost of 13 extra parameters seems like a bad trade. To quantify this, once again the AIC is useful.Fig. 3.3 compares the nominal and

effective fits (top and middle panels). The likelihood of the effective model is about $1e^{-12}$ for the same dataset as Fig. 3.1, hence the nominal model is preferable.

- Linear regression (also known as Ordinary Least Squares—OLS) might not be the preferred method. If the goal is to predict MT from ID_e, then OLS is the adapted tool, but if the goal is to estimate the slope between MT and ID_e, techniques such as Major Axis, or Standardized Major Axis are more appropriate (see e.g. [156, 153]). An example of the OLS versus the SMA fit is given in the lower panel of Fig. 3.3. With this particular example, differences in estimated a and b are minor.

3.4 Measuring Performance: Throughput

Originally, Fitts was measuring the capacity of the human motor system [38]; this notion is still used, although often in a non-technical sense. When evaluating whether one input device is superior to another, HCI researchers compute the throughput of the device. Sometimes, they will also refer to (interaction) bandwidth [170] or information transfer rate [85]. All of these notions can be used interchangeably; ultimately, the goal of throughput and its synonyms is to compute a rate of information in bits/second.

There is no theoretical definition for throughput, only various operational formulas (e.g. [170] lists three). The two most common formulations are

$$\mathrm{TP} = 1/b, \text{ and} \tag{3.22}$$

$$\overline{\mathrm{TP}} = \mathbb{E}\left[\frac{\mathrm{ID_e}}{\overline{\mathrm{MT}}}\right], \tag{3.23}$$

where $\mathbb{E}$ abusively[4] refers to averaging.

The confusion of having two widespread formulas is probably due to Fitts [170]. In his seminal paper [38] Fitts defined an Index of Performance (IP) as

$$\mathrm{IP} = \mathrm{ID}/\mathrm{MT}, \tag{3.24}$$

but later [42] redefined IP as

$$\mathrm{IP} = 1/b. \tag{3.25}$$

Evidently, (3.25) is equivalent to (3.22), and (3.23) is just an average formulation of (3.24), with the correction for accuracy given in 3.3.

Both formulations for throughput compress a and b in a single metric. In the case of TP (3.22), the intercept a is simply discarded. The definition is in line with Fitts'

[4] Its computation depends on the experiment e.g., whether or not there are multiple participants or just one. The idea is that for each condition, the effective index is computed and divided by the average movement time $\overline{\mathrm{MT}}$. These ratios are then averaged over all blocks (i.e. for all conditions, participants etc.).

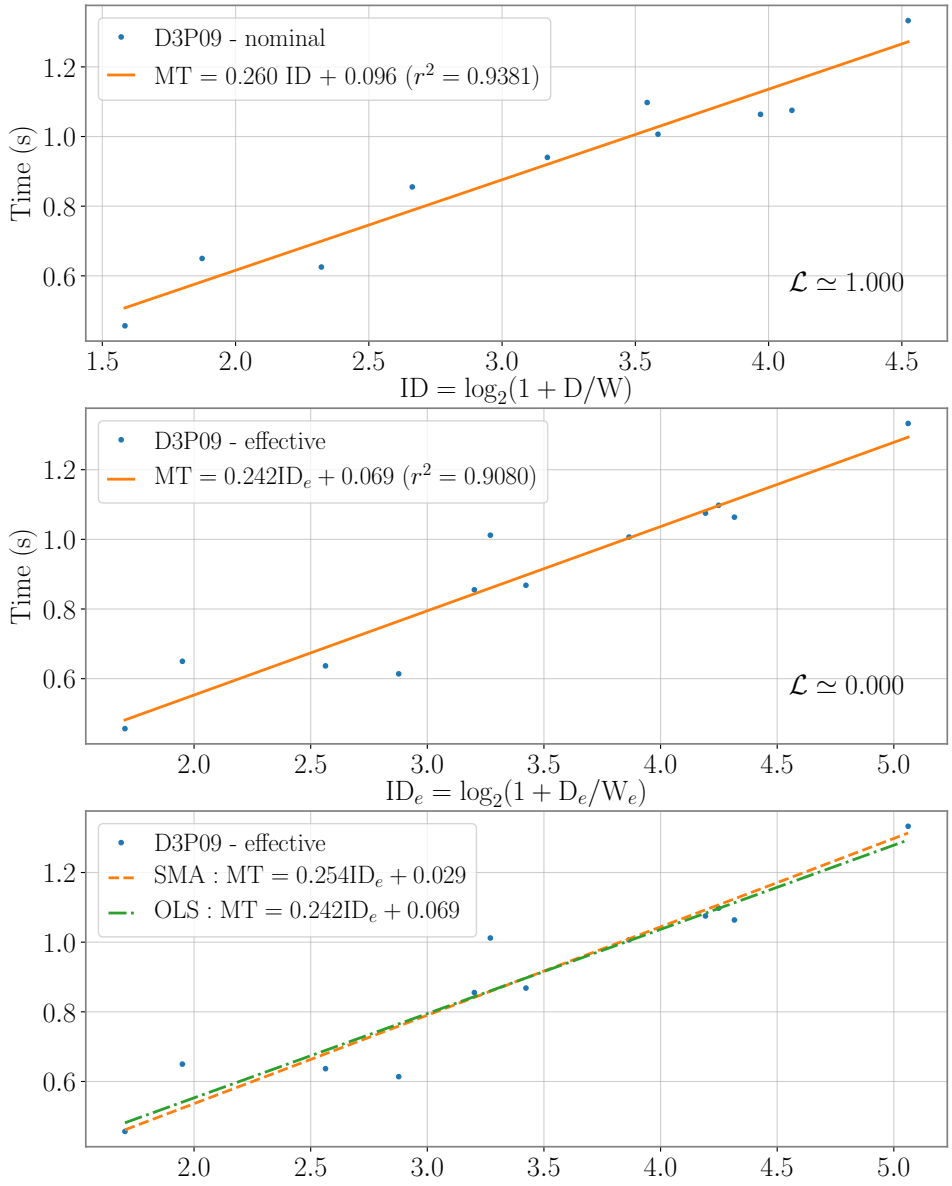

Figure 3.3: Top Panel: Nominal pointing data from the JGP dataset (see Chapter 7) and corresponding linear regression. Middle Panel: Effective pointing data and corresponding linear regression. Bottom Panel: Effect of SMA versus OLS on a and b from the effective formulation (3.19). $\mathcal{L}$ is the likelihood of the model as defined in (3.15).

theory (1.15), which does not predict an intercept. For common devices such as the mouse, TP is usually about 10 bits/second.

On can argue that just leaving a out is unwarranted, after all part of the time spent in pointing is due to a, and sometimes a can be significantly different from 0 (e.g. in the semminal study by Card et al. [11], the intercept is more than a second). This is precisely the idea behind $\overline{\mathrm{TP}}$ (3.23). One problem that results from using $\overline{\mathrm{TP}}$ is that the throughput is then dependent on the range of ID investigated [170]. Indeed,

$$\overline{\mathrm{TP}}^{-1} = \frac{a + b\,\mathrm{ID}_\mathrm{e} + \varepsilon}{\mathrm{ID}_\mathrm{e}} = b + \frac{a + \varepsilon}{\mathrm{ID}_\mathrm{e}}, \tag{3.26}$$

where ε is the residual in the linear regression. The consequence on the evaluated throughput can be dramatic. By reanalizing pointing data for a mouse [11], Zhai [170] shows that removing a single condition can shift the $\overline{\mathrm{TP}}$ value up to about 20%, while it has barely any effect on TP.

Computing $\overline{\mathrm{TP}}$ brings a second difficulty, as it requires averaging outputs of non linear functions. The so-called Jensen's inequality [113] then asserts that the order in which data is averaged matters; from a personal simulation I estimated that in the most extreme cases $\overline{\mathrm{TP}}$ may vary by a factor of 2, and that in realistic cases $\overline{\mathrm{TP}}$ could vary by about one bits/second.

On one hand, TP discards valuable information by not accounting for a, while on the other hand $\overline{\mathrm{TP}}$ is substantially contaminated by experimental design and data processing. We are forced to conclude that no adequate throughput measure exists. Equation (3.26) reveals that in the case of $\overline{\mathrm{TP}}$ the main problem lies with the intercept a. It thus appears crucial to provide a good interpretation of a, to decide whether or not it should be included in the measure of throughput. This will be achieved in Chapters 6 and 9.

3.5 Discussion

At this point, I suspect the reader might be somewhat perplexed. Although I have emphasized the success of Fitts' law in Chapter 1 in HCI and in experimental psychology, I have also pointed to the weak theoretical formulation — but supposedly its empirical regularity was a good enough reason to uphold it. In this section however, I have also pointed towards several empirical deficiencies or difficulties associated with the use of Fitts' law in HCI. These problems are not new, in fact the problems of throughput and of nominal versus effective measures have been discussed thoroughly by Zhai in the Special Issue of the International Journal of Human-Computer Studies celebrating the fiftieth anniversary of Fitts' law [64]. Judging by the popularity of Fitts' law since this 2004 issue (335 results for the keyword "Fitts' law" in the CHI' proceedings alone, between 2004 and 2018), the law has withstood the criticism. I believe there are two reasons to this.

First, not all researchers are actually aware of some of the pitfalls of Fitts' law. The law, taught in most HCI courses, seems immune to any suspicion. The standardization

of Fitts' law [1] probably adds to the credibility of the law.

Second, the problems described in this chapter are especially harmful when comparing two independent studies. For example, everyone will agree that a mouse is preferable to a joystick, yet the throughput of the joystick according to Card et al. [11] is about 5 bits/second, whereas the throughput for a mouse according to Mackenzie et al. [98] is about 4 bits/second. However within the same study, often comparisons do hold; Card et al. also estimated the throughput for a mouse to be about 10 bits/second, which does show superiority of the mouse over the joystick. Hence, comparisons when needed can be performed — they just come at the cost of running ever more empirical studies. How problematic is this ? For the HCI community taken as a whole, this certainly means a slower progress and more opacity. For the HCI researcher individually, this might give an easy opportunity to publish quick results.

3.6 Appendix

Computation 1. Consider the random variable for the endpoint location Y, such that $Y \sim \mathcal{N}(0, \sigma^2)$ and a target of width W. The event $|Y| > \mathrm{W}/2$ defines a miss. Width W and miss rate $\varepsilon > 0$ are related through the following one-to-one relation

$$\varepsilon = 1 - 2\int_0^{\frac{\mathrm{W}}{2}} \frac{1}{\sqrt{2\pi}\sigma} \exp\Big(-\frac{t^2}{2\sigma^2}\Big) dt = 1 - \operatorname{erf}\Big(\frac{\mathrm{W}}{2\sqrt{2}\sigma}\Big) \tag{3.27}$$

or

$$\mathrm{W} = 2\sqrt{2}\sigma \operatorname{erf}^{-1}(1-\varepsilon), \tag{3.28}$$

where erf is the Gaussian error function

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\frac{t^2}{2}} dt. \tag{3.29}$$

These formulas are consistent with recommendations in [142]: Taking $\mathrm{W} = 4.133\ \sigma$, we find

$$\varepsilon = 1 - \operatorname{erf}\Big(\frac{2.066}{\sqrt{2}}\Big) = 3.88\%. \tag{3.30}$$

This 3.88% is usually rounded off to 4% [142]. The multiplicative constant α such that $\mathrm{W}_e = \alpha \mathrm{W}$, where $\mathrm{W}_e = 4.133\ \sigma$ is the width such that the error rate is 3.88%, is given by

$$\alpha = \frac{\mathrm{W}_e}{\mathrm{W}} = \frac{4.133\ \sigma}{2\sqrt{2}\ \sigma \operatorname{erf}^{-1}(1-\varepsilon)} = \frac{2.066}{\sqrt{2} \operatorname{erf}^{-1}(1-\varepsilon)}. \tag{3.31}$$

so that W_e is given by the following formula:

$$\mathrm{W}_e = \alpha \mathrm{W} = \frac{2.066}{\sqrt{2} \operatorname{erf}^{-1}(1-\varepsilon)} \mathrm{W}. \tag{3.32}$$

To compare this to recommendations in [142], consider the z-score z, related to the area under a $\mathcal{N}(0,1)$ distribution by

$$\int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = x \iff 1 - Q(z) = x \iff z = Q^{-1}(1-x) \tag{3.33}$$

where Q is Marcum's Q-function and Q^{-1} the inverse Q function. The inverse Q-function can be easily related to the inverse error function

$$Q^{-1}(y) = \sqrt{2}\,\mathrm{erf}^{-1}(1-2y), \tag{3.34}$$

and by replacing y by $1-x$, we find that

$$z = \sqrt{2}\,\mathrm{erf}^{-1}(2x-1). \tag{3.35}$$

Replacing x by $1 - \frac{\varepsilon}{2}$ gives the final result:

$$W_e = W \frac{2.066}{z(1-\frac{\varepsilon}{2})} = \frac{2.066}{\sqrt{2}\,\mathrm{erf}^{-1}(1-\varepsilon)} W. \tag{3.36}$$

Part II

Information-Theoretic Models for Voluntary Aimed Movement

Chapter 4

A Basic Source Model: Aiming is Choosing

As pointed out in the introduction, Fitts first reasoned by analogy with source coding results. In this very short chapter, we derive the most popular formulations of ID, namely (3.3), (3.5) and (3.7) from a simple source coding argument.

Fitts' paradigm is at first sight not suited to an information-theoretic analysis. In a reciprocal task (see Fig. 1.3), there are two identical targets, while in the discrete task, there is only a single target. Entropy of the set of targets is then respectively smaller than 1, and null, which is not very useful. However, understanding that a rectangle is indeed a target is a perceptual process, one that the motor system does not necessarily comprehend; and indeed the motor system can position the limb between targets. Hence, the free space between the targets is also a potential target for the motor system.

Fig. 4.1 displays the problem from the motor system's point of view. Knowing the full extent of movement D and the required precision W, the problem of aiming simply reduces to the problem of choosing the correct target among all possible targets that make up the free space. Throughout this thesis, I refer to this principle as "aiming is choosing".

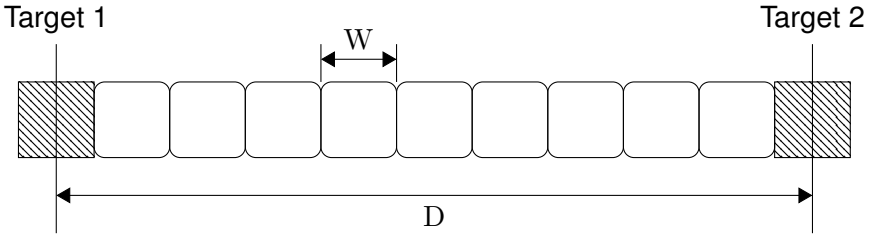

Figure 4.1: Placing targets and identifying D and W in the context of the MacKenzie formulation.

Let $\mathcal{S}$ be the set of all targets for the motor system; we can look to assign probabil-

ities to each one of these targets. We do not assume that the motor system expects a point in space to be more likely than another, hence, the probability that one target is indeed the right target is *a priori* only determined by its size when compared to the total space. As all targets are the same size, the probability p of one target being the right target is simply

$$p = \frac{\mathrm{W}}{\mathrm{D}+\mathrm{W}}. \tag{4.1}$$

If we assume for simplicity that $(\mathrm{D}+\mathrm{W})/\mathrm{W}$ is a round number, the number of targets N that fit inside $\mathrm{D}+\mathrm{W}$ is exactly

$$N = (\mathrm{D}+\mathrm{W})/\mathrm{W}. \tag{4.2}$$

We can now compute the entropy of the set of targets $H(\mathcal{S})$:

$$H(\mathcal{S}) = -\sum_{i=1}^{N} p\log_2 p = -Np\log_2 p = \log_2 N \tag{4.3}$$

$$= \log_2\bigl(1+\mathrm{D}/\mathrm{W}\bigr), \tag{4.4}$$

which exactly equals the Mackenzie formulation of ID (3.7).

The index of difficulty is here computed as a source entropy—there is no information transmission. The aiming task is simply identified to the creation of target identifiers using the "aiming is choosing" rationale. Since the uniform distribution is the one that *maximizes* entropy [18] it provides the least upper bound on the entropy for *any* target probability distribution. The resulting entropy $H(\mathcal{S})$ is thus the number of bits required to identify the target position without any prior knowledge whatsoever, and ID arises as a measure of the uncertainty associated with the task of choosing one target. The more potential targets (the higher the ratio D/W), the more difficult the pointing task.

We can obtain other formulation for ID by changing the layouts of targets. Indeed, what can be considered free space is quite subjective, and two other possibilities are explored in Figs. 4.2, 4.3 and 4.4.

The calculation of (4.4) applied to the layout of Fig. 4.2 results in an entropy $H(\mathcal{S}) = \log_2\bigl(1/2+\mathrm{D}/\mathrm{W}\bigr)$ that matches Welfords' formulation (3.5). The layout comes from considering the discrete aiming task which has a single target, rather than the reciprocal task above.

The calculation of (4.4) applied to the layout of Fig. 4.3 results in an entropy $H(\mathcal{S}) = \log_2 \mathrm{D}/\mathrm{W}$ that matches Crossman' formulation (3.4). The layout also comes from considering the discrete aiming task, when considering that the largest extent of the movement is D.

Finally, we can also match Fitts' formulation (3.3) using the layout of Fig. 4.4. In this case, choice of direction is also considered, on top of choices of amplitudes. From the middle starting point, we have Crossman's case to the right, and to the left its symmetric; the total free space is thus $2\mathrm{D}$, for $H(\mathcal{S}) = \log_2 \frac{2\mathrm{D}}{\mathrm{W}} = 1+\log_2 \mathrm{D}/\mathrm{W}$. Equivalently, to

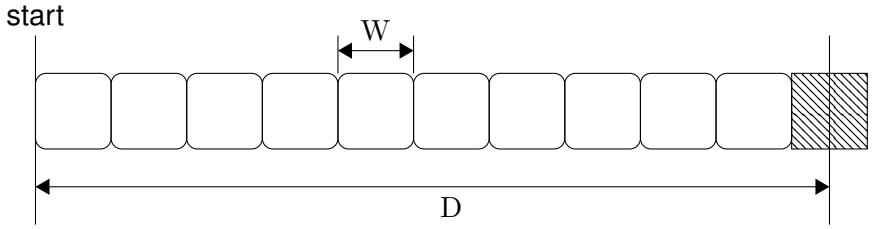

Figure 4.2: Placing targets and identifying D and W in the context of the Welford formulation.

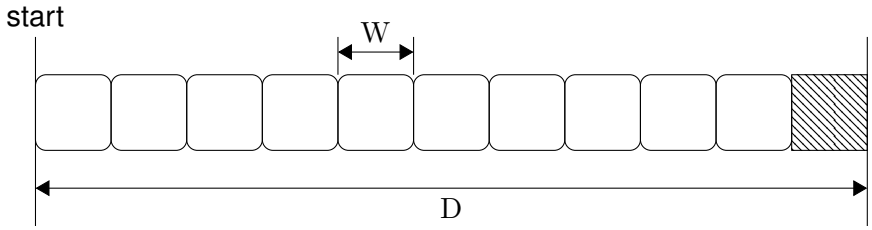

Figure 4.3: Placing targets and identifying D and W in the context of the Crossman formulation.

obtain Fitts' formulation, one specifies first the direction, and then the amplitude as in Crossman's case. Because there are only two equiprobable directions, and the direction is independent from the amplitude, the total entropies is simply the sum of the entropy related to the choice of direction (1 bit) and the Crossman entropy ($H(\mathcal{S}) = \log_2 \mathrm{D}/\mathrm{W}$).

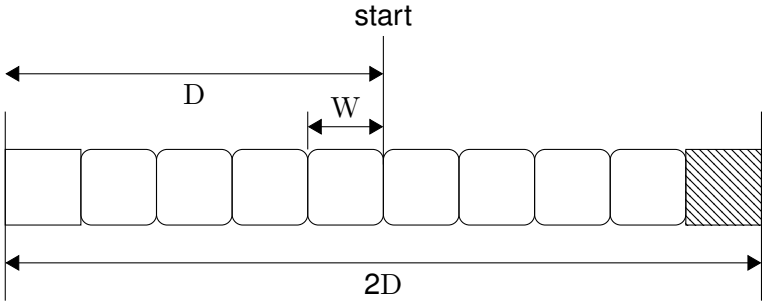

Figure 4.4: Placing targets and identifying D and W in the context of the Fitts formulation.

Fitts' law then results by considering that MT is linearly related to $H(\mathcal{S})$. All 4 layouts are permissible, and any one of them can be advocated based on the experimental design e.g., whether or not the task is discrete or reciprocal. "Aiming is choosing" will thus not guide us to the "correct" formulation, but as discussed in 3.2, this is not an important point.

What is more interesting is that Fitts [37] explicitly used the term information "gen-

eration" rather than transmission and made movement time depend upon the index of difficulty. This is consistent with his assumption that the ID should serve to characterize target entropy—a *source* coding rate in Shannon's sense, and also consistent with this "aiming is choosing" rationale. It is somewhat surprising that to justify the same index in his famous article published one year later, Fitts [38] referred to Theorem 17—a *channel* coding rate in Shannon's sense, which is completely unrelated to source coding.

Chapter 5

Feedforward Transmission Model with Bounded Noise: A Formal Information Theoretic Transmission Scheme (FITTS 1)

Fitts based his rationale for the description of voluntary movements on information-theoretic concepts; unfortunately we have seen that the resulting Fitts' law was merely based on a vague analogy with Shannon's Theorem 17. The most successful attempt at explaining the law, Meyer et al.'s SOS model, does not build on Shannon's information theory, but rather on the optimization of a two component model. The goal of this chapter is to push Fitts' idea as far as possible and provide a derivation for Fitts' law based on the limited capacity of a transmission channel. We will see that the analogy can be justified; even though in the end this first model will teach us little about motor control.

5.1 Voluntary Movement as a Transmission Problem

If we wish to consider movement as a transmission problem, then we must necessarily identify the components of Shannon's communication paradigm Fig. . In particular, we have to identify a source, a transmitter, a channel, a receiver and a destination, as well as to determine what the relevant signal is. A single channel hypothesis with a structure for the chain of mechanisms involved in sensory-motor performance was previously discussed by Welford [157], but on a generic level. More recently Zhai et al. [172, p. 106, Fig. 2.1] proposed a model for stroke gestures, in which the human intention forms the source of the communication system.

5.1.1 Black-box Model

A complete model for pointing that would be obtained by considering all the components of the CNS that take part in generating a movement is necessarily very elaborate. A much simpler description is attempted at the top of Fig. 5.1. The user intention, which serves as the source outputs some message. The form of this message is hard to fathom, and is probably best seen as an abstract representation located somewhere in the brain. It is then converted through various parts of the brain to a set of electrical signals through the operation we call movement mapping. This transforms the abstract intention into a tangible signal. This signal is sent through the nervous system, where it gets corrupted by (neural) noise. The signal is received by the motor organs and the ensuing limb extremity is moved appropriately so that the movement ends somewhere in space, close enough to the intended area. If the participant is aiming towards a specific target, it is either hit or not. Although all these functions are described in very vague terms, there is a great difficulty in proposing a precise characterization of each block of this model.

What we show in this chapter is that the model can be reduced to a much simpler one, see the bottom of Fig. 5.1, where no encoding or decoding procedures are required. This leads to somewhat of a black-box model, where only input and output are considered and all the difficulties of modeling movement are taken into account by a well chosen noise Z.

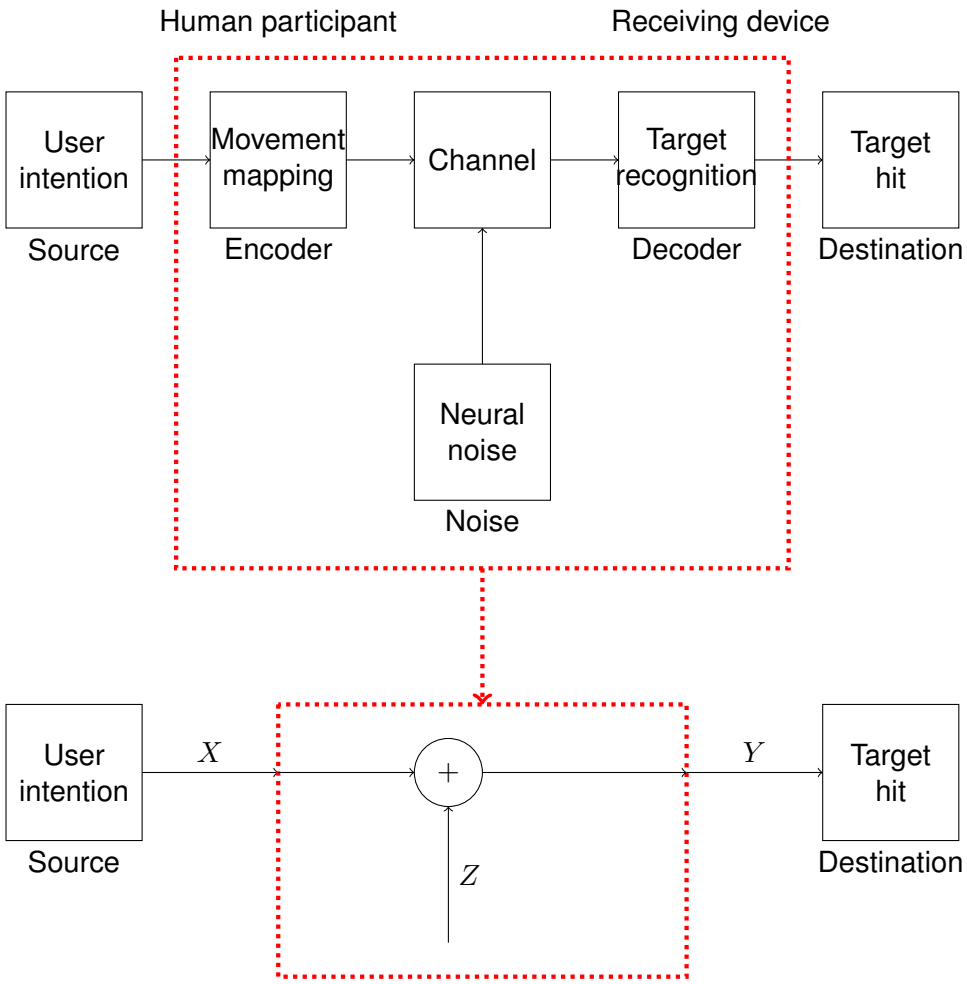

Figure 5.1: The human motor system as a communication system.

5.1.2 Model Description

We now get into the details of the blocks of the model at the top of Fig. 5.1.

Source The information source we consider is the user's intention. An immediate difficulty is that in Fitts' paradigm, there is a single target to hit, so that there is no a priori information to transmit to the receiver. This same problem was solved in 4 with the "aiming is choosing" rationale. We consider the targets of Figure 4.1; we have that the message X takes value in the set $\{-\frac{D}{2}, -\frac{D}{2}+W, \ldots, \frac{D}{2}-W, \frac{D}{2}\}$, see Fig. 5.2(b), where all possible values of X are marked as black bullet points. As seen in 4, the entropy of the source then reduces to the MacKenzie index of difficulty: $H(X) = \mathrm{ID} = \log_2(1+D/W)$: The smaller the targets, the higher the source entropy. Smaller targets that are located further away are more difficult to represent.

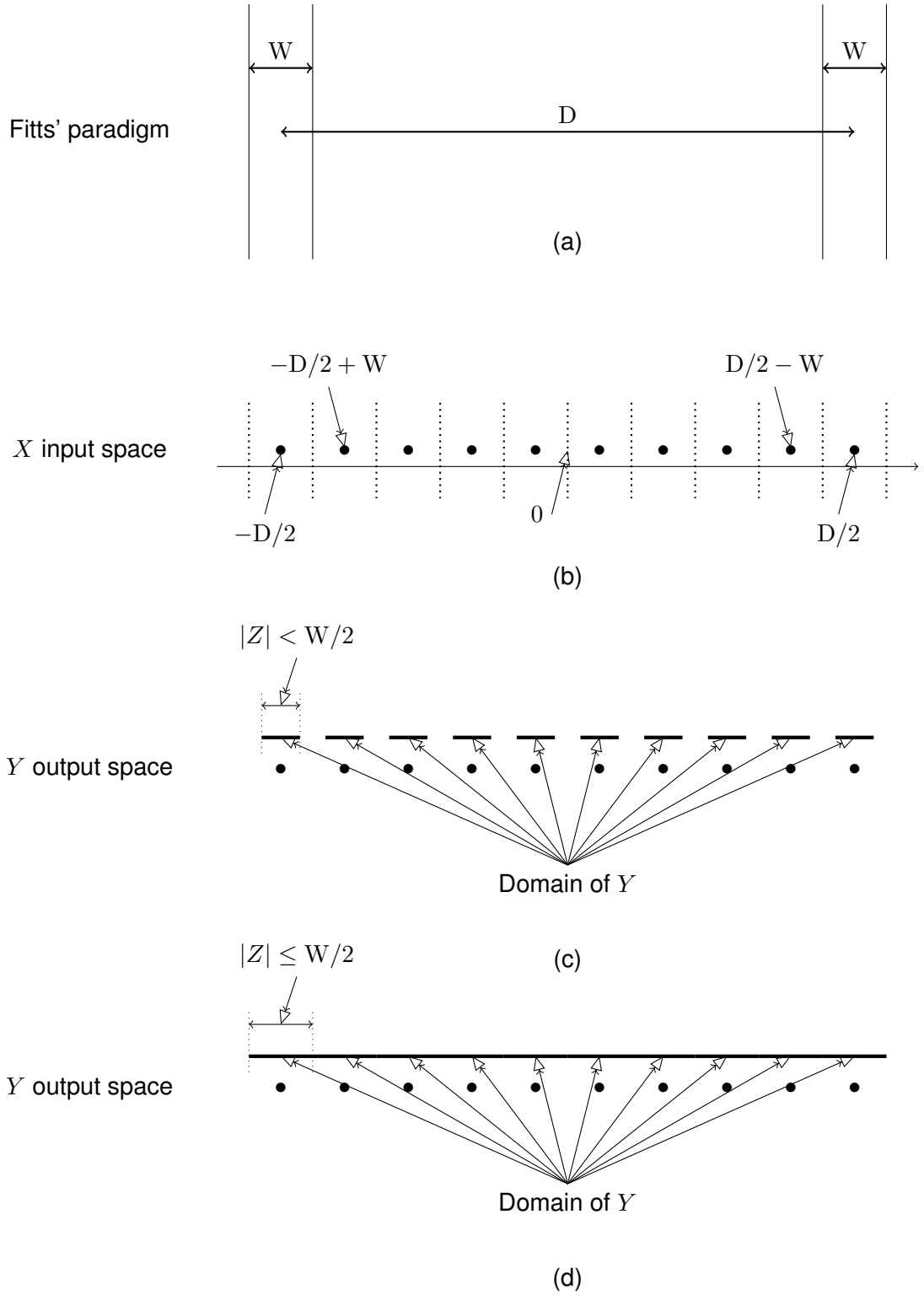

Figure 5.2: (a) Fitts' paradigm for the reciprocal task. (b) Discrete input space (Domain of X). Each dot is a possible value for X and corresponds to the center of each target. (c) Domain of Y in the case where the amplitude of the noise $|Z|$ is less than W, made of the reunion of all the segments. (d) Domain of Y when $|Z| \leq \mathrm{W}$. The domain of Y is the interval $[-\frac{\mathrm{D}+\mathrm{W}}{2}; \frac{\mathrm{D}+\mathrm{W}}{2}]$.

Channel The message produced under the intention of the participant is encoded and sent through the noisy channel. The noise in the channel is presumably a reflection of the imperfection of neural and musculo-skeletal mechanisms, and should ultimately model movement end-point variability.

Destination At the output of the channel, the signal Y can attain many different values; in fact the support of Y i.e., the set of values that Y can assume is made of the reunion of all the intervals that are centered around the different values of X and are as wide as the amplitude of Z, see Fig. 5.2 (c) and (d).

5.1.3 Aiming Without Misses

If we want the target to be aimed at to become effectively hit, then the noise must have an absolute amplitude less than $\mathrm{W}/2$, i.e., $|Z| \leq \mathrm{W}/2$. The fact that aiming is successfully carried out thus entails an amplitude constraint on the channel, rather than the usual power constraint that is often assumed in digital communications.

The receiver simply checks which target has been attained, by evaluating to which interval Y belongs. It may be the participant herself/himself, through a visual check (which suggests the possibility of some feedback, see next section for a treatment of feedback). It may also be a computer if we are considering pointing with a cursor.

The amplitude constraint on the noise ensures that the right target is always hit, which ensures errorless communication. In summary, our model for the aiming task is comprised of a source that corresponds to the "aiming is choosing" paradigm, and a limited-amplitude channel that allows the receiver to ensure that the target is never missed. This limited-amplitude channel is further described next.

5.2 Fitts' law and the Capacity of the Uniform Channel

In this section we derive Fitts' law from the computation of the capacity of a particular transmission scheme which we call the uniform channel.

5.2.1 Uniform Channel Versus the Gaussian Channel

As discussed previously, there is a mismatch between Fitts' law and Shannon's capacity formula: in Shannon's formula, the term inside the logarithm is a ratio of powers, whereas in Fitts' formula the ratio is one of amplitudes.

The fact that Shannon's formula displays powers can be sketched as follows:

1. Following the Channel Coding Theorem, the capacity of a channel where the input is X and the output is Y is found when maximizing $I(X;Y)$ over all input distributions. This entails finding not only the right set for X, but also which probability to assign to each element of the set.

2. If the noise Z is *independent* from the input i.e., the voltage of the noise does not depend on the voltage from the input, and *additive*, i.e., the voltage of the noise adds up to the voltage of the input to produce the voltage of the output, then the mutual information can be developed as:

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(X+Z|X) = H(Y) - H(Z). \quad (5.1)$$

   Thus, for an independent and additive noise, because the entropy of the noise is what it is whatever the input, maximizing mutual information over the input is equivalent to simply maximizing the entropy of the output $H(Y)$. Additive and independent noise are very commonly assumed in digital communications.

3. Gaussian noise is the most natural model for noise in a point-to-point communication channel. Indeed, we can assume that the noise voltage is due to the sum of many random and independent components, which asymptotically follows a Gaussian distribution. Furthermore, the Gaussian distribution allows tractable analytical studies, which as well explains its wide usage.

4. In practice a receiver is conceived with some nominal *electrical* power, which should only rarely and shortly be exceeded; it shouldn't draw too much power, or it may, for example, overheat. This electrical power is proportional to the *mathematical* power and as a result, there is a power constraint in the sense of signal processing, on Y. Because the power of the noise is what it is, the constraint on output power is also a constraint on input power. If N is the power of the noise, and Q the power constraint on Y, then there is a power constraint $P = Q - N$ on X. Generally, because we are optimizing $I(X;Y)$ over X, we assign the constraint on the input.

5. Maximizing $H(Y)$ under a power constraint is achieved when Y assumes a Gaussian distribution.

6. The entropy of a Gaussian distribution can easily be expressed as the logarithm of the powers of the signal. For example, for the case of the noise $Z \sim \mathcal{N}(0, N)$

$$H(Z) = \frac{1}{2}\log\big(2\pi e N\big) \quad (5.2)$$

7. Noise being independent from the input, the power of the output is simply obtained as the sum between the power of the input P and the power of the noise N. Because Y assumes a Gaussian distribution, we have that

$$H(Y) = \frac{1}{2}\log\big(2\pi e(P+N)\big) \quad (5.3)$$

8. C is obtained as the difference between $H(Y)$ and $H(Z)$:

$$C = H(Y) - H(Z) \quad (5.4)$$

$$= \frac{1}{2}\log\frac{P+N}{N} \quad (5.5)$$

In our aiming problem, the constraint we put is on the amplitude of the signals. The range of the signal should not be greater than D, while the range of the noise should not be greater than W. The departure with Shannon's scheme is thus at item 4. Under an amplitude constraint, the distribution that maximizes the noise is the uniform distribution; therefore one might expect considering the uniform distribution rather than the Gaussian distribution to prove useful. We will show next that this is indeed the good idea.

5.2.2 Capacity of the Uniform Channel

The results of this subsection are adapted from [125], which compares "Hartley's rule" and the Shannon capacity and concludes that Shannon's capacity formula holds for a whole range of channels, from the uniform to the Gaussian. Hartley's rule is a formula which is very similar with the MacKenzie ID, as it also involves the logarithm of a ratio of amplitudes instead of powers.

We first sum up the few steps that lead to the capacity formula for the uniform channel, in the spirit of the previous subsection.

1. Following the Channel Coding Theorem, the capacity of a channel where the input is X and the output is Y is found when maximizing $I(X;Y)$ over all input distributions. This entails finding not only the right set for X, but also which probability to assign to each element of the set.

2. If the noise Z is *independent* from the input i.e., the voltage of the noise does not depend on the voltage from the input, and *additive*, i.e., the voltage of the noise adds up to the voltage of the input to produce the voltage of the output, then the mutual information can be developed like so:

$$I(X;Y) = H(Y) - H(Y|X) = H(Y) - H(X+Z|X) = H(Y) - H(Z). \quad (5.6)$$

   Thus, for an independent and additive noise, because the entropy of the noise is what it is whatever the input, maximizing mutual information over the input is equivalent to simply maximizing the entropy of the output $H(Y)$.

3. The aiming task imposes an amplitude constraint on the noise and on the input signal, namely $|Z| \leq \mathrm{W}/2$ and $|X| \leq \mathrm{D}/2$. This yields an amplitude constraint on Y, namely $|Y| \leq \frac{\mathrm{D}+\mathrm{W}}{2}$.

4. Maximizing $H(Y)$ under an amplitude constraint is achieved when Y assumes a uniform distribution.

5. There is no obvious choice for the distribution of the noise, as we cannot access the physical channel for measurements. However there are two reasons that lead us to consider the uniform distribution. First, it allows a simple analytical study. Second, the uniform noise maximizes entropy under amplitude constraint, so that uniform noise is essentially a *worst-case* scenario, and is thus the one that we

should assume according to the maximum entropy principle. It also mimics the scenario in Shannon's formula, where the noise distribution maximized its entropy.

6. The entropy of a uniformly distributed signal can easily be expressed as the logarithm of the range of the signal. For example, for the case of the noise $Z \in [-\mathrm{W}/2; \mathrm{W}/2]$

$$H(Z) = \log \mathrm{W}. \tag{5.7}$$

7. Similarly, the entropy of the output is

$$H(Y) = \log\bigl(\mathrm{D} + \mathrm{W}\bigr) \tag{5.8}$$

8. C is obtained as the difference between $H(Y)$ and $H(Z)$:

$$C = H(Y) - H(Z) \tag{5.9}$$

$$= \log \frac{\mathrm{D} + \mathrm{W}}{\mathrm{W}} = \log\bigl(1 + \mathrm{D}/\mathrm{W}\bigr). \tag{5.10}$$

These steps can be formalized. We first define the uniform channel.

Definition 5.2.1 (Uniform Channel)**.** A uniform channel has the following properties:

1. an additive and independent noise:

$$Y = X + Z; \tag{5.11}$$

2. an amplitude constraint on noise;

3. a uniformly distributed noise.

Calculating the capacity of this uniform channel yields MacKenzie's ID.

Theorem 5.2.1. *The capacity C' of the uniform channel under the amplitude constraint $|X| < \frac{D}{2}$ is given by the following expression:*

$$C' = \log\Bigl(1 + \frac{\mathrm{D}}{\mathrm{W}}\Bigr) \qquad \textit{bpcu}.$$

The proof uses the following lemma, and is given in the Appendix.

Lemma 5.2.1. *Consider the uniform channel. If the output is uniformly distributed in $[-(D+W)/2, (D+W)/2]$ and the noise is uniformly distributed in $[-W/2, W/2]$, then the input must be a uniform discrete random variable in the set $\{-D/2, -D/2 + W, \ldots, D/2 - W, D/2\}$.*

This Lemma can be understood quite easily by looking at Fig. 5.2. In the two lower panels, black lines represent the probability density function of Y. In panel (c), the probability density function is piecewise uniform, because the amplitude of the noise is too low to fill the entire interval. In panel (d), the amplitude of the noise is just the right amplitude to fill the entire interval, yielding a uniform distribution for Y. Should the amplitude of the noise be larger, there would be overlaps and the probability density function of Y would not be uniform any more.

Instead of modifying the range of the noise, we can equivalently modify the spacing between each value that X can assume. It is immediate to see that values of X should be spaced W apart to produce a uniform output Y. The formal proof can be adapted from [125].

Lemma 5.2.1 indicates what the input should look like to obtain a uniform output Y, and it turns out this distribution is exactly the target distribution of Fig. 4.1 corresponding to the Mackenzie formulation. Not only does the MacKenzie ID match the entropy of the target distribution, it also matches the capacity of the channel used in modeling the aiming task. An important result of the proof is that the capacity-achieving input distribution corresponds exactly to the uniform channel's input, meaning that no channel coding is required: Sending messages from the source directly over the channel is optimal. What then distinguishes good from poor performances is bandwidth only. Theorem 5.2.1 also implies that:

$$C' = \max_{p(x)} \big(H(X) - H(X|Y) \big) = H(X), \tag{5.12}$$

meaning that in the optimal scheme, no information is lost in the channel since $H(X|Y) = 0$.

5.2.3 Uniform Noise

The main reason to opt for the uniform distribution is from a mathematical standpoint (see [124] and [139] for a similar use); and in fact with this distribution we have obtained a simple derivation of the MacKenzie ID. However, most practitioners would probably object to a uniform noise (even though at least one experiment [140] reports almost uniform variability for wrist movements); in fact in most cases the shape of the distribution of endpoints is positively or negatively skewed, or symmetric, depending on e.g. the average speed reached during the movements [71]. In practice, it is very common to assume a Gaussian distribution for endpoints [21, 142] or the less precise notion of bell shaped curve [142]. The next theorem gives reason to consider the uniform noise nevertheless.

Theorem 5.2.2. *Consider C'', the capacity of any channel that is amplitude limited. Then we have that C', the capacity of the channel with the same amplitude limitation but with uniform noise, is always smaller than C'':*

$$C'' \geq C'. \tag{5.13}$$

C' is thus a lower bound of the actual capacity C'' with amplitude limitation and arbitrary shape. Furthermore, with a noise that has an amplitude bounded Gaussian distribution with $W/\sqrt{2\pi e} = \sigma$, we have that

$$C' \leq C'' \leq C' + 0.2. \tag{5.14}$$

The proof is given in the Appendix. In practice[1], when assuming an amplitude bounded Gaussian distribution for noise with $W/\sqrt{2\pi e} = \sigma$ (this corresponds to the 4% error case in [142]), we have that $C' \leq C'' \leq C' + 0.2$, which makes for a reasonably tight bound.

5.2.4 Analogy with Shannon's Capacity Formula

Since the MacKenzie index involves a ratio of amplitudes D/W rather than a ratio of powers P/N, it is appropriate to compute it in terms of powers to further the analogy. The surprising result is that the ID is mathematically equivalent to Shannon's Capacity. This is expressed in the next theorem.

Theorem 5.2.3. *Let $C = (1/2)\log_2(1 + P/N)$ denote the Shannon's capacity and $C' = \log_2(1 + \mathrm{D}/\mathrm{W})$ denote the capacity for the uniform channel, then:*

$$C' = C \qquad \textit{bpcu}.$$

The formal proof is in the Appendix, but this coincidence can be explained quite easily when noticing that $1 + \mathrm{SNR} = (P + N)/N$ is the ratio between the power of the output Y over the power of the noise Z. In our case, both output and noise are uniformly distributed, the power is proportional to the square of the range of the distribution, so that

$$\frac{P + N}{N} = \Big(\frac{\mathrm{D} + \mathrm{W}}{\mathrm{W}}\Big)^2, \tag{5.15}$$

Taking the logarithm gives $C' = C$. For this particular channel the index of difficulty and the Shannon capacity truly coincide, legitimizing the analogy with Shannon's Theorem 17. We now formulate a proper analogy from Shannon's capacity formula rearranged in the following manner:

$$C = B_W \cdot \log_2(1 + \mathrm{SNR}) = 2B_W \cdot \frac{1}{2}\log_2(1 + \mathrm{SNR}) \qquad \text{bit/s.} \tag{5.16}$$

From theorems 5.2.1 and 5.2.3, we can identify $1/2\log_2(1 + \mathrm{SNR})$ with MacKenzie's index of difficulty $\mathrm{ID} = \log_2(1 + D/W)$. Also, by virtue of the Shannon-Nyquist sampling theorem, $2B_W$ refers to the maximum number of samples that are sent per second, which can be identified to $\frac{1}{\mathrm{MT}}$ as we effectively perform one movement during MT seconds. We thus obtain Fitts' Formula (1.15)

$$\mathrm{MT} = \frac{1}{C}\mathrm{ID}, \tag{5.17}$$

[1] A Gaussian distribution of movement endpoints is often assumed with $W \simeq 4.133\sigma$ [142], see 3.3

without intercept. Interestingly, Fitts did not refer to an intercept in his 1954 article. He introduced it later to make the model more flexible for experimental data. The interpretation of the intercept has been debated many times (e.g. [171, 142, 65], see also 3.4).

5.3 Taking Target Misses into account

In the previous section, we considered a uniform distribution for noise, and showed that, as long as the participant did not miss the target, it provided a lower bound on the achievable rate. In practical scenarios, the difference between considering the uniform distribution and the amplitude-bounded Gaussian distribution is relatively small. The previous model is therefore sufficient as a description in a paradigm that does not allow mistakes, such as Fitts' pin and disc transfer experiments [38, Experiments 2 and 3]. However, in the majority of Fitts' law experiments target misses do occur, and so an extension of the model is needed; the goal of this section is therefore to consider the general case where the participant might miss the target.

5.3.1 Handling Misses

In a Fitts' law experiment the outcome of a pointing act can be either measured as an error, i.e., a distance from end-point to target center, or categorized in an all-or-none way as a hit versus a miss (see [55]). Information theory offers a useful distinction between transmission *errors* (the received symbol is wrong) and *erasures* (the received symbol is empty). This distinction seems to have escaped attention so far in HCI, where it has been a solid tradition, since MacKenzie [94], to measure movement endpoints from the center of the target and, assuming that the distributions of these measures is normal, to compute an effective index of difficulty $\mathrm{ID_e}$.

The goal of a Fitts' law experiment being to observe and study the speed-accuracy tradeoff, the choice of the metrics used to measure *speed* and *accuracy* is critical. While there has been unanimous agreement in the literature that movement time provides a satisfactory measure of speed (see next section), the measurement of accuracy has been controversial from the outset [21]. In HCI, it is customary to proceed to an adjustment for target misses [142, 1] through the effective index of difficulty $\mathrm{ID_e}$. Unfortunately, as shown below, this standardized method is not rigorous.

There are three different ways of handling mistakes:

- *Ignoring the mistakes*. Fitts, who did not measure actual amplitudes, classified the movements in a dichotomous way as hits and misses. Although he did tabulate the error rates he obtained in his stylus-pointing experiments, he felt in a position to leave them aside because of the "small incidence" of target misses [38, p. 265].
- *Taking the error rate into account*. To our knowledge, Crossman [21] was the first to try to incorporate the error rate information into his ID measure, leveraging the standard Gaussian distribution model.

- *Taking the spread of endpoints into account*. This is the standardized way of measuring accuracy in Fitts' law [142, 1] trough $\mathrm{ID_e}$, see 3.3. Recourse to the standard deviation as a measure of accuracy has the implication that the magnitude of the metrical error (the distance from target center) matters in the upcoming analysis: Regardless of whether the outcome is a hit or a miss, the farther the endpoint from target center, the worse the performance. It also implies that there is equivalence between two movements hitting the target if and only if they end up at exactly the same distance from the center of the target.

The ISO standard, and the Fitts' law literature in general, treats pointing mistakes as *errors*, by referring to the standard deviation of the endpoints distribution—either by direct estimation or through a calculation from error rates. Thus in the *error* concept, the accuracy depends on the (continuous) distance between the movement endpoint and the target center. This approach is not quite consistent with the all-or-none logic of Fitts' experimental paradigm. In an experiment that asks participants to *hit* the target 96 percent (or so) of the time, all movements that end up inside the W interval should be recognized as equivalent from the point of view of accuracy. Importantly, that equivalence is true in real-world interfaces: What matters is not precisely *where* the click takes place, but rather *whether or not* the click falls in the intended area. In the information-theoretic vocabulary, this type of alteration is known as an *erasure* [55].

Not only did we previously underline the theoretical and empirical difficulties of using $\mathrm{ID_e}$ in 3.3, there is also a conceptual mismatch between the standardized measurement of accuracy and the reality of the pointing task in both controlled experiments and real-world target acquisition tasks.

5.3.2 A Compliant Index of Difficulty: $\mathrm{ID}(\varepsilon)$

We now propose a new effective index, which we denote $\mathrm{ID}(\varepsilon)$, that is compliant with Fitts' experimental design, does not rely on any endpoint distribution hypothesis, and is justified theoretically as a channel capacity, through an extension to the model presented Section 5.2.

As noted above, treating target misses as transmission errors is not adapted to Fitts' paradigm—these events should rather be viewed as erasures. In fact, the design of the experiment entails a binary decision: The movement either finishes inside the target (a hit) or outside of it (a miss). This is consistent with the instruction "try to hit the target" as opposed to "try to hit the center of the target". We now extend the model that does not allow or account for mistakes of the previous section with a channel which allows erasures.

Consider a channel that oscillates randomly between a good (G) state and a bad (B) state, with probability ε of being in state B and probability $1-\varepsilon$ of being in state G. When the channel is in its good state, it corresponds to the channel of capacity $\log\left(1+\frac{\mathrm{D}}{\mathrm{W}}\right)$ that we derived in Section 5.2, which we will refer to as the *Fitts channel*. However, when the channel is in its bad state it can only produce erasures—we call it

an *erasure channel*. In Information Theory this configuration (Figure 5.3) is known as a *compound* channel [49].

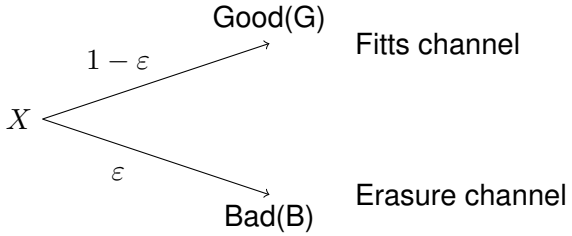

Figure 5.3: Compound channel for an aiming task with target misses.

Let us now evaluate the Shannon capacity of this compound channel. This will serve as a common ground to compare the performance of different participants operating at different accuracy levels (with different values of ε). The channel capacity corresponds to the maximum transmission rate that the participants would have achieved with an arbitrarily small error rate (refer to the Channel coding theorem of Chapter 1). We thus adjust the rate, to obtain the rate that the participants would have had, had they never missed the target. Shannon's capacity of the compound channel of Figure 5.3 is given by the following theorem.

Theorem 5.3.1. *Consider a compound channel as in Figure 5.3, with probability ε of being in state B and probability $1-\varepsilon$ of being in state G. The capacity of such a channel is given by*

$$C = (1-\varepsilon)\log_2\Big(1+\frac{\mathrm{D}}{\mathrm{W}}\Big).$$

As expected, the obtained capacity is lower than the capacity $\log_2\big(1+\frac{D}{W}\big)$ that would have been achieved with 100% hitting success ($\varepsilon = 0$).

For completeness, we provide the proof of this theorem (this is a known result; see e.g., the classic textbook [18], where this proof is given as an exercise)

Proof. Since the only way to produce an erasure symbol is for the channel to be in state B, we have $I(X;Y) = I(X;(Y,E))$. This can be expanded as

$$\begin{aligned} I(X;(Y,E)) &= \mathbb{P}(E=G)I(X;Y|E=G) + \mathbb{P}(E=B)I(X;Y|E=B) && (5.18)\\ &= (1-\varepsilon)\, I(X;Y|E=G) + \varepsilon I(X;Y|E=B). && (5.19)\end{aligned}$$

Here $I(X;Y|E=G)$ is the mutual information computed for the uniform channel, and $I(X;Y|E=B) = 0$ because if the channel is in bad state, only an erasure can come out of the channel. Therefore the distribution that maximizes the mutual information for the compound channel is the same as that which maximizes mutual information for the uniform channel. □

The participant is effectively time sharing both channels. With Fitts' channel, the transmitted information is $\log_2\Big(1+\frac{\mathrm{D}}{\mathrm{W}}\Big)$ bits and with the erasure channel the transmitted information is 0 bit, so that, on average, $C=(1-\varepsilon)\times\log_2\Big(1+\frac{\mathrm{D}}{\mathrm{W}}\Big)+\varepsilon\times 0$. In line with Fitts' parallel between capacity and ID, our new effective index is

$$\mathrm{ID}(\varepsilon)=(1-\varepsilon)\log_2\Big(1+\frac{D}{W}\Big) \tag{5.20}$$

where ε is no other than the traditional 'error rate' more cautiously designated here as the percentage of target misses.

5.3.3 Comparison between $\mathrm{ID_e}$ and $\mathrm{ID}(\varepsilon)$

As noted before, the behavior of the standardized $\mathrm{ID_e}$ for vanishing error rates is problematic. As shown in Computation 1 in 3.6, asymptotically what matters is the behavior of the inverse Gauss error function (or equivalently the behavior of the z-score, as $z(x)=\sqrt{2}\,\mathrm{erf}^{-1}\,(2x-1)$.

It is well known that $\mathrm{erf}^{-1}(1-\varepsilon)$ tends to $+\infty$ as ε vanishes, because the Gaussian distribution is strictly positive over $\mathbb{R}$. Correspondingly, we should find that $\mathrm{ID_e}$ diverges:

$$\lim_{\varepsilon\to 0}\mathrm{ID_e}=\infty. \tag{5.21}$$

However, as explained previously, it is usually recommended to enforce an arbitrary threshold ($\varepsilon=0.0049\%$) below which $\mathrm{ID_e}$ is kept constant.

Therefore, if we evaluate the "bounded version" of $\mathrm{ID_e}$, we obtain

$$\begin{aligned}\lim_{\varepsilon\to 0}\mathrm{ID_e}&=\log_2\Big(1+\frac{D}{0.5089W}\Big)\simeq\log_2\Big(1+\frac{2D}{W}\Big)\\&=1+\log_2\Big(\frac{1}{2}+\frac{D}{W}\Big),\end{aligned}$$

which is equivalent to the Welford index of difficulty [157], by direct application of corollary 3.2.1. The arbitrary choice to bound the index at the 0.0049% rate results in the index coincidentally tending to the Welford ID, not the MacKenzie ID. Thus there is no continuity[2] as epsilon approaches zero for $\mathrm{ID_e}$. In contrast, $\mathrm{ID}(\varepsilon)$ *does* have the property of continuity towards zero since obviously $\mathrm{ID}(0)=\mathrm{ID}$.

Figure 5.4 shows the two indices $\mathrm{ID}(\varepsilon)$, $\mathrm{ID_e}$ as well as the unbounded u-$\mathrm{ID_e}$ (for which the 0.0049% distinction is not made) for $D/W=15$ as a function of ε in the interval $[0-1]$. The difference $\mathrm{ID_e}-\mathrm{ID}(\varepsilon)$ between $\mathrm{ID_e}$ and $\mathrm{ID}(\varepsilon)$ is lowest around $\varepsilon=0.1$. With higher values of ε, the difference increases but such high errors rates are not common. However, for very small values of ε, $\mathrm{ID}(\varepsilon)$ can be up to 1 bit smaller than $\mathrm{ID_e}$. Thus the difference between $\mathrm{ID}(\varepsilon)$ and $\mathrm{ID_e}$ can be non-negligible for very careful participants or in conditions with a high emphasis on accuracy, or, in real-world interfaces, in conditions where the cost of any pointing mistake is deterring.

[2]Not in the sense of a mathematical continuity, but in the sense that there is switch from one index to an other.

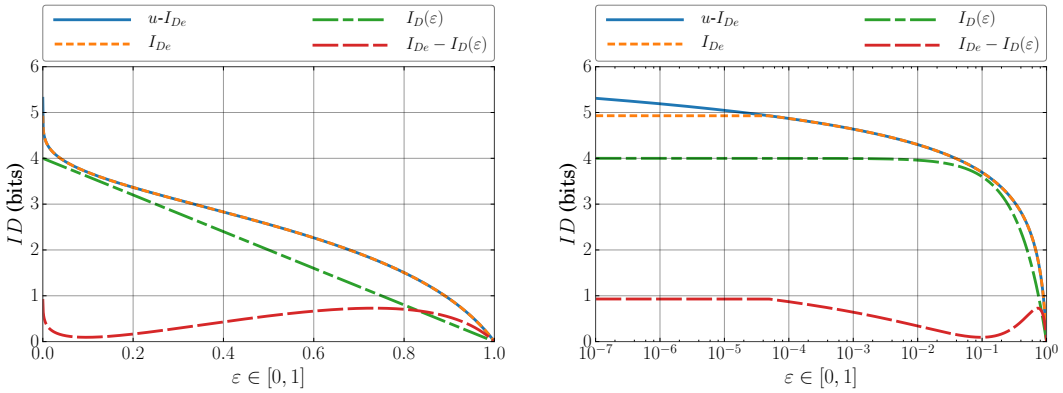

Figure 5.4: Comparison of $\mathrm{ID}(\varepsilon)$ and $\mathrm{ID_e}$ for erasure rate in $[0, 1]$, for $\frac{\mathrm{D}}{\mathrm{W}} = 15$ ($\mathrm{ID} = 4$). u-$\mathrm{ID_e}$ refers to the "unbounded" $\mathrm{ID_e}$ where the 0.0049% threshold is not considered. The scale is lin-lin in the left panel and lin-log in the right.

5.4 Performance Fronts for Fitts' law

Fitts' law has always been thought of as a law of average performance 3.1.1. Although the notation does not make it explicit, MT, the dependent variable of (1.17), typically denotes the *mean* of samples of movement time measures. Soukoreff and MacKenzie [142] state that " Each condition must be presented [...] many [...] times, so that the *central tendency* of each subject's performance [...] can be ascertained".[3] Researchers have "agreed to disagree" on many issues of Fitts' law, e.g. on which formulation to use for the index of difficulty, on how to account for errors, and on the meaning of the intercept, see Chapter 3. However, most if not all Fitts' law students have agreed on using the *linear regression* to describe the relation between ID and MT. That technique provides both an estimate of parameters a (intercept) and b (slope) and a measure of goodness of fit, through the *r*-squared coefficient, in a very simple and rapid manner. Likewise, throughput is usually computed as a mean of means [142] (but see also 3.4). Its identification to a channel capacity seems, therefore, problematic since the channel capacity concept has nothing at all to do with average information transmission performance: Only the *best* transmission schemes are capacity achieving.

In the remainder of this chapter, we build on Guiard and colleagues [66, 67] claim that *only* the best movement times can serve to infer Fitts' law.

5.4.1 Fitts' Law as a Performance Limit

We see two reasons for viewing Fitts' law as a performance limit rather than a law of average performance.

[3] Emphasis added.

1. Fitts' information-theoretic rationale for aiming considers the transmission of information about the target through a human motor channel, and as we have shown Fitts' law can be derived by computing the *capacity* of this channel, which is a theoretical upper bound—the maximum amount of information that can be transmitted reliably—and which is accordingly calculated as an extremum through the Channel Coding Theorem—the maximum of mutual information over all input distributions. Thus, only movements that maximize transmitted information can be expected to be described by Fitts' law, i.e. those movements that for a fixed ID achieve the lowest MT, or conversely those that for a fixed MT achieve the highest ID;

2. Guiard and Rioul [67] have shown that three paradigms, the time minimisation paradigm of Fitts [38], the spread minimisation paradigm of Schmidt et al. [129], and the dual minimisation paradigm of Guiard et al. [66] can receive a unified account provided that the participants are assumed to invest less than 100 percent of their resource in their performance.[4] Accordingly, only the best performing samples should be expected to describe the speed-accuracy tradeoff, and Guiard and Rioul then successfully merge the linear law of Schmidt et al. [129] and the logarithmic description of Fitts' law as different regions of the same speed-accuracy tradeoff function.

To understand the constraints on movement, one should consider the movements that are most constrained: One can only hope to model what can be modeled. In the real world, movements are weakly constrained, if they are at all. One rarely tries to point as fast and as accurately as possible. Even in a controlled experiment, the participants' attention fluctuates. Predicting everyday unconstrained performance requires a complex model, where Fitts' law probably has little guidance to offer.

As we demonstrate now, an approach that considers only the best movements, which we call the front of performance, can reveal Fitts' law even in real-world data. In this chapter, an example that will help bring the ideas across is provided.

5.4.2 A Field Study Example

The data come from a pointing study run "in the wild" by Chapuis et al. [14] (CBB dataset, see Chapter 7). While delivering very large data sets, field experiments (as opposed to controlled experiments) provide a beneficial magnification of the fact that not all resources are invested by participants for each movement.

For several months Chapuis et al. [14] unobtrusively logged cursor motion from several participants using their own hardware. The authors were able to identify offline the start and end of movements as well as the relevant target characteristics, for several

[4] Not only does the less-than-total resource investment assumption match common sense, it matches the information-theoretic concept of capacity. The capacity is reached at the limit of a (perfectly) optimal coding scheme, channel bandwidth being exploited in full. Anything less will give lower transmitted information.

hundreds of thousands of click-terminated movements. With this information, one can then represent the movements in an MT versus ID graph, as normally done in a controlled Fitts law study. To compute task difficulty in the 2D space of computer screens they chose [97]

$$\mathrm{ID} = \log_2\Big(1 + \frac{D}{\min(H, W)}\Big), \tag{5.22}$$

where H and W are the height and width of the target, respectively. Whenever an item was clicked , it was considered the target, meaning the rate of target misses was 0 percent and hence $\mathrm{ID}(\varepsilon) = \mathrm{ID}(0) = \mathrm{ID}$.

Figure 5.5 shows the data from one representative participant (P3) of the CBB dataset (see Chapter 7). The ID axis is truncated at 6 bits because beyond that level of difficulty the density of data points dropped significantly. Obviously, the data obtained with no speeding instructions (and no experimenter to recall them) exhibits a huge amount of stochastic variability along both dimensions of the plot. While in the X dimension, most ID values fell in the range from 0.5 to 6 bits (presumably a reflection of the geometric composition of the graphical user interface), the variability along the Y dimension is extremely high. Judging by linear regression on this raw data, we find that movement time and the index of difficulty are essentially *uncorrelated* since the r-squared coefficient is very close to 0 ($r^2 = 0.034$). Thus, at first sight, this data fails to confirm Fitts' law[5], but it is important to realize that this first impression is quite false.

In the bottom panel of Figure 5.5 which ignores all MT data above 4 s and thus zooms-in on the Y-axis towards the bottom of the plot, one can distinctly see that the bottom edge of the cloud of data points does not touch the X axis. Rather, in the downward direction, the density drops sharply: No matter the ID region considered, the distribution of performance measures has an unending *tail* above and what we call a *front* below, the latter being very steep in comparison with the former. The unending tail is understandable as "it is always possible to do worse" [67]. In contrast, the movement time cannot be reduced below a certain strictly positive critical value which accurately defines the front.

A closer look at the bottom panel of Figure 5.5 reveals that the bottom edge of the scatter plot is approximately *linear*. This linear edge is what justifies Fitts' law. In other words, the empirical regularity in Fitts' law is, in essence, a front of performance, a lower bound that cannot be passed by human performance. Such a front of performance is observable in data from the field study of Chapuis et al. [14] because unsupervised everyday pointing does offer, albeit in a minority of cases, opportunities to perform with high levels of speed and accuracy. The difference between a field and a controlled experiment is thus one of degree, not of nature. Experimenters have recourse to pressurizing speed/accuracy instructions simply to get rid of endless, uninformative, tails in their distributions of MT measures.

[5]Even though, inevitably with more 90,000 pairs of values, a coefficient of correlation $r = 0.184 (r^2 = 0.034)$ is significant at p < .0001.

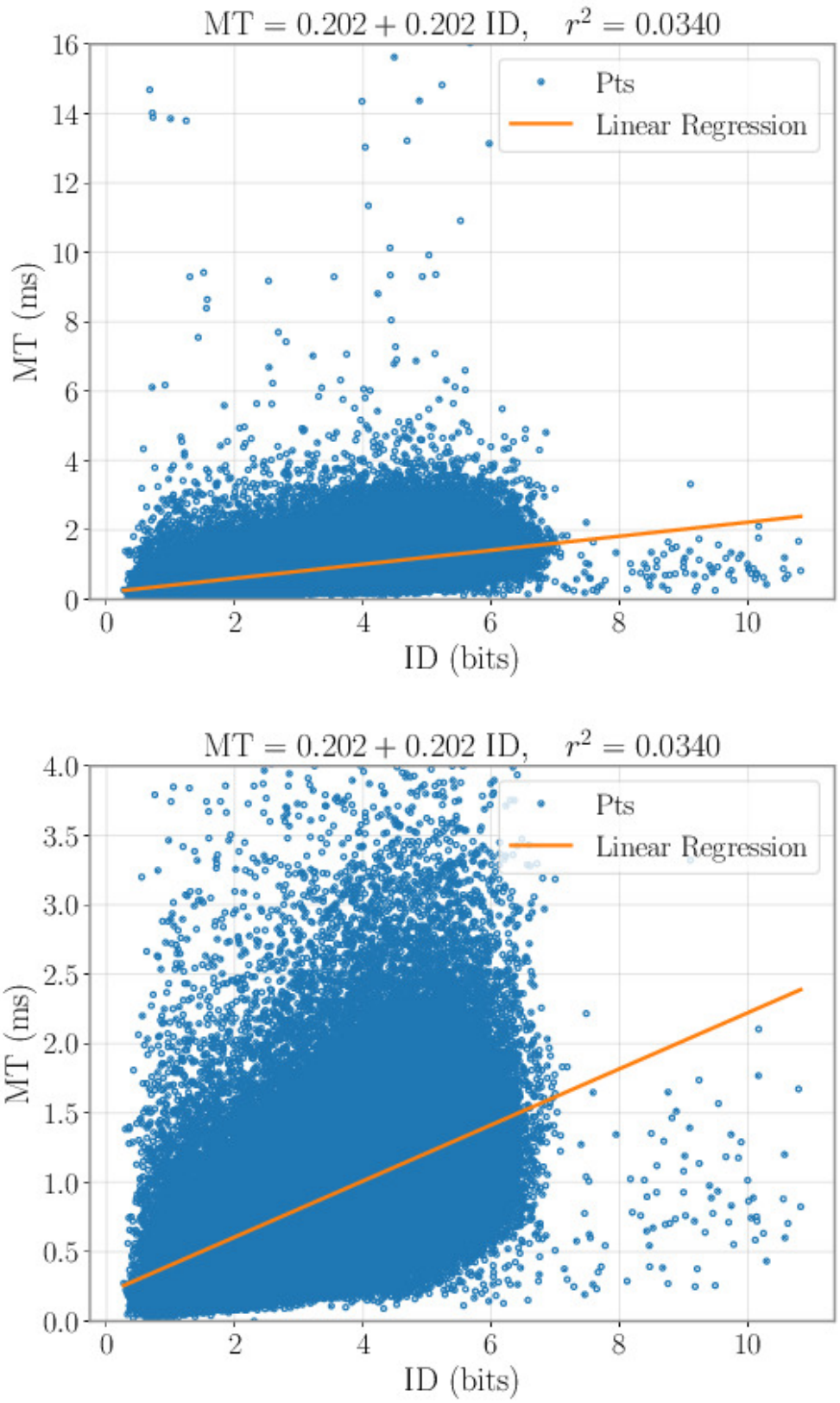

Figure 5.5: Movement time as a function of task difficulty in one representative participant of the CBB dataset. Shown are over 90,000 individual movement measures. *Left*: MT up to 16 s. *Right*: MT up to 4 s. Cut-offs are here arbitrary but necessary as some movement times lasted several seconds.

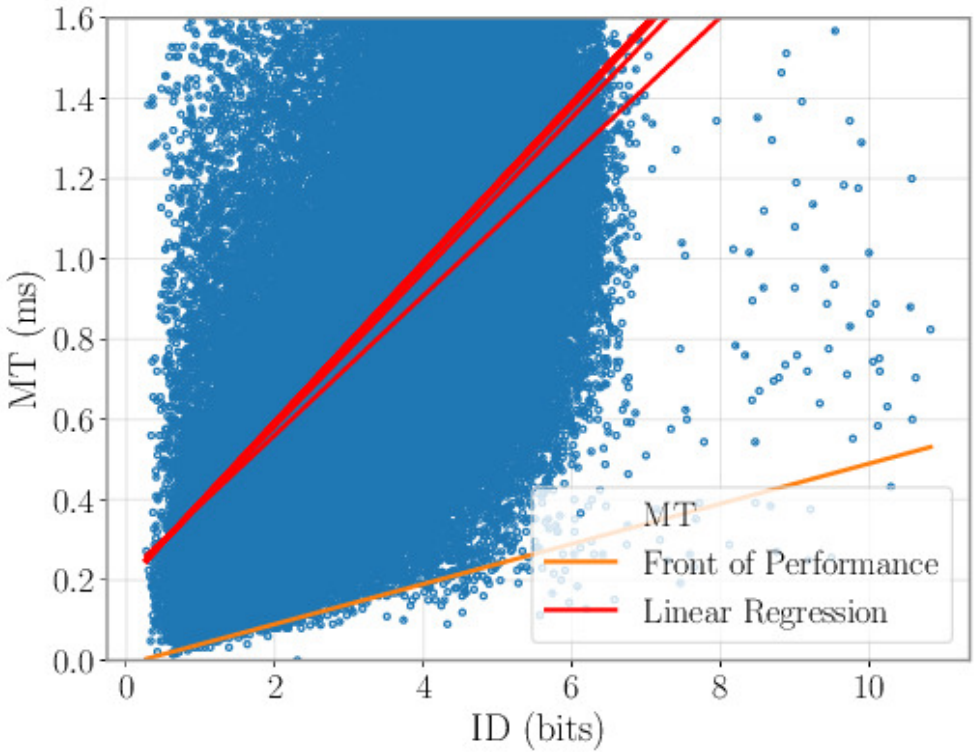

Figure 5.6: Same data as in Figure 5.5, with the Y-axis cut at 1.6 s. Shown in red are linear fits from usual linear regression, using a number of different thresholds (from 2 s to 15 s) for the exclusion of outliers, as well as an estimation of the front of performance (in orange).

Figure 5.6 shows the same plot with the Y axis zoomed-in further so that the range of MT measures approaches that commonly obtained in a typical controlled experiment. Even though the front edge is incomparably sharper than the tail edge, the zoomed-in view of Figure 5.6 reveals a number of presumable *fast* outliers. Many reasons may explain why a small proportion of data points "cross" the frontier, seemingly violating the theoretical lower bound. Some data points may just correspond to unreasonably fast but lucky movements, others to failures of the analysis software, which may have wrongly classified as target-directed movements which terminated with accidental clicks. Yet another possibility is that targets lying at the edge of the screen can be aimed at with a purely ballistic throw of the cursor which will remain on that edge. An empirical scatter plot will never exhibit a perfectly neat front of performance, and so an estimation procedure is still needed to actually estimate the front.

Figure 5.6 shows an ideal front fit, in orange at the bottom edge of the plot. The obtained line is (ideally) independent of *slow* outliers. In contrast, linear regression lines obtained with different threshold levels $[2$ s$, 3$ s$, ..., 9$ s$, 15$ s$]$ for outlier rejection (in red in Figure 5.6) show that they are highly dependent on the threshold level.

Thus, an interesting characteristic of the front of performance approach is that it dispenses us with the difficult task of handling *slow* outliers, whose removal requires arbitrary choices. For example, some experimenters remove values k standard deviations away from the sample mean, k being typically chosen between 2 and 3. Some simply trim the data, by removing all samples above a certain limit, say $\mathrm{MT} > 2s$. One issue here is that the tolerance for outliers is variable across the ID scale. As illustrated Figure 5.6, the fit computed by linear regression highly depends on the arbitrary choice of tolerance. In contrast, the front of performance *by definition* does not depend on slow outliers at all, and in this sense it is far more robust.

Of course, the orange line is quite different from the red lines in Figure 5.6: Characterizing Fitts' law by best rather than average performances is not a minor adjustment, especially in "wild" experiments. Even though experimenters do their best to reduce the inherent variability of human aimed movement, a typical sample of measures exhibits quite large dispersion. The common practice of considering averages per block, rather than raw measures, reduces this dispersion artificially. This practice does not eliminate the fact that because of movement variability, the quantitative difference between average fit and best-performance fit is substantial.

A method to fit a straight line to the bottom of the edge of the scatter plot, robust enough to accommodate the imperfectness of the front, is given in 8.3.2.

5.5 Appendix

Proof of Theorem 5.2.1. We can now expand the mutual information $I(X;Y)$ as difference of differential entropies:

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X), && (5.23)\\ &= H(Y) - H(X+Z|X), && (5.24)\\ &= H(Y) - H(Z) && (5.25)\\ &= H(Y) - \log_2(W), && (5.26) \end{aligned}$$

where (5.23) is by definition of mutual information, (5.24) is by additivity of the channel, (5.25) is by independence of X and Z, and (5.26) is from the computation of differential entropy for a continuous uniform random variable. Maximizing $I(X;Y)$ is thus equivalent to maximizing $H(Y)$. Because $|X| \leq \frac{\mathrm{D}}{2}$ and $|Z| \leq \frac{\mathrm{W}}{2}$, we have that $|Y| = |X+Z| \leq |X| + |Z| \leq \frac{\mathrm{D}+\mathrm{W}}{2}$ by the triangular inequality. The maximum

$$C' = \max_{|X| \leq \frac{\mathrm{D}}{2}} I(X,Y) = \max_{|Y| \leq \frac{\mathrm{D}+\mathrm{W}}{2}} H(Y) - \log \mathrm{W}.$$

will be attained when Y is uniformly distributed in $[-(\mathrm{D}+\mathrm{W})/2, (\mathrm{D}+\mathrm{W})/2]$ and from Lemma 5.2.1, this is obtained when X is discrete uniform in the set: $\{-\mathrm{D}/2, -\mathrm{D}/2 + \mathrm{W}, \ldots, \mathrm{D}/2 - \mathrm{W}, \mathrm{D}/2\}$. It follows that

$$C' = \log_2(\mathrm{D}+\mathrm{W}) - \log_2(\mathrm{W}) = \log_2\Big(1 + \frac{\mathrm{D}}{\mathrm{W}}\Big) \qquad \text{bpcu.}$$

as claimed. □

Proof of Theorem 5.2.2. According to Theorem 3 in [125], the capacity C'' associated with an arbitrary additive noise Z of limited amplitude such that $|Z| \leq W/2$ satisfies:

$$\log_2(1 + D/W) \leq C'' \leq \log_2(1 + D/W) + log_2(\alpha),$$

where $\alpha = W/2^{h(Z)}$. Let us present an example using a noise with a zero-mean Gaussian with standard deviation σ, whose probability density function is

$$p(z) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{z^2}{2\sigma^2}).$$

The corresponding amplitude-limited noise Z has density $\tilde{p}(z)$ which vanishes for $|z| > W/2$ and is otherwise equal to

$$\tilde{p}(z) = \frac{p(z)}{c}$$

where from Computation 1 in 3.6, $c = c(W, \sigma) = \operatorname{erf}(\frac{W}{2\sqrt{2}\sigma})$. Its entropy $\tilde{h}(z)$ is then

$$\begin{aligned}\tilde{h}(z) &= -\int \tilde{p}(z)\log_2 \tilde{p}(z)\,dz = -\int_{-W/2}^{W/2} c^{-1}p(z)\log_2\big(c^{-1}p(z)\big)\,dz \\ &= \frac{1}{2}\log_2(2\pi\sigma^2c^2) + c^{-1}\int_{-W/2}^{W/2}\frac{p(z)z^2}{2\sigma^2}(\log_2 e)\,dz\end{aligned}$$

Integrating by parts we obtain

$$\tilde{h}(z) = \frac{1}{2}\log_2(2\pi\sigma^2c^2) + (\frac{1}{2}\log_2 e)\left[1 - \frac{W}{c\sqrt{2\pi}\sigma}\exp -\frac{W^2}{8\sigma^2}\right]$$

We observe the following:

- for $\sigma << W$, since $\lim_{x\to\infty}\operatorname{erf}(x) = 1$ and the negative exponential dominates the outermost right term of the entropy, we find that $\tilde{h}(z) \simeq \frac{1}{2}\log_2(2\pi e\sigma^2)$, which is the entropy of the Gaussian distribution. This was expected since when σ is small enough, we have $p(z) \simeq \tilde{p}(z)$;
- for $\sigma >> W$, we use the fact that for small x, $\operatorname{erf}(x) \simeq \frac{2x}{\sqrt{\pi}}$ at first order in x, giving, after some computations, $\tilde{h}(z) \simeq \log_2(W)$. Hence $\tilde{p}(z)$ behaves, as expected, like the uniform distribution when σ is large enough.

The value of $\log_2(\alpha)$ is evaluated for practical ratios $0.1 \le W/\sigma \le 10$ in Fig. 5.7. In particular, we get that $\log\alpha = 0.203$ for $\mathrm{W}/\sigma = 4.133$, which we can reasonably round off to 0.2. □

Proof of Theorem 5.2.3. Let M be the cardinality of the set $\{-\mathrm{D}/2, -\mathrm{D}/2+\mathrm{W}, \ldots, \mathrm{D}/2-\mathrm{W}, \mathrm{D}/2\}$:

$$M = 1 + \frac{\mathrm{D}}{\mathrm{W}}.$$

The channel input's average power is:

$$P = \mathbb{E}(X^2) = \frac{1}{M}\sum_{k=0}^{M-1}\Big(\frac{M-1}{2} - k\Big)^2\mathrm{W}^2 = \frac{1}{M}\,2\mathrm{W}^2\sum_{k=0}^{\frac{M-1}{2}}k^2 = \frac{M^2-1}{12}\mathrm{W}^2$$

where we have used the well-known formula for the sum of consecutive squares. The noise power N of the uniformly distributed distribution in $[-\mathrm{W}/2, \mathrm{W}/2]$ is

$$N = \frac{\mathrm{W}^2}{12}.$$

It follows that

$$C' = \log_2 M = \frac{1}{2}\log_2 M^2 = \frac{1}{2}\log_2(1 + M^2 - 1) = \frac{1}{2}\log_2\Big(1 + \frac{P}{N}\Big) = C.$$

as claimed. □

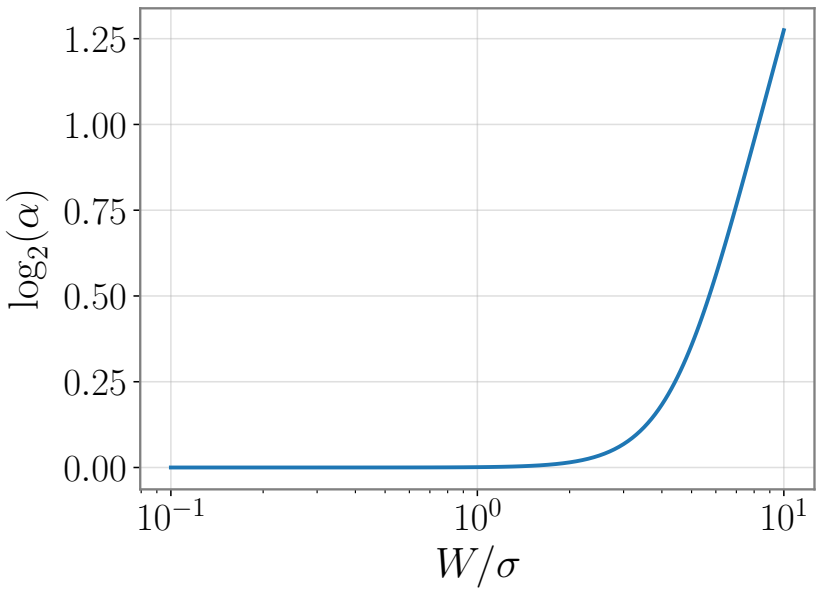

Figure 5.7: $\log_2(\alpha)$ evaluated for $0.1 \leq W/\sigma \leq 10$.

Chapter 6

Feedback Transmission Model with Gaussian Noise: A Feedback Information Theoretic Transmission Scheme (FITTS 2)

FITTS 1 unambiguously links ID to a channel capacity in a well-defined, simple aiming scheme where input output and noise are given an interpretation. However, FITTS 1 is a black-box model for endpoints; it is only concerned with MT (i.e. neglects most of the trajectory to focus only on the latter's extremities), and does not take advantage of the empirical findings in neuro-science, experimental psychology and behavioral science, while remaining quite abstract.

The model discussed in this chapter, FITTS 2, perfects these shortcomings, and is voluntarily driven by the observations of Chapter 2: that human produced movements display great variability, and can be produced by feedback or feedforward, as well as intermittent or continuous control.

Obviously, accommodating for these observations while retaining a simple model is far from trivial. The trick that we use in this chapter is to consider a large set of trajectories and track how its variance evolve over time, through our so-called PVP's. Considering many trajectories has the benefit of averaging off all intermittent control into a seemingly continuous process. We then provide a model with both a feedforward and feedback channel which explains the PVP's.

6.1 Positional Variance Profiles (PVP)

Trajectory variability has "surprisingly received little attention from researchers" [146]. Most studies investigating Fitts' law have often been limited to the measure of the spread of endpoints [21, 142], yet "spatial error is a dynamic feature of every moment, and not only a static feature of endpoint accuracy" [151]. The models that we reviewed

in 2.2 that handle full trajectories account for single movements [102, 117]. As, we have shown in 2.1.3, considering single movements comes with practical difficulties, such as parsing movements, in particular because movements can appear to be continuously or intermittently controlled. In this section, we consider what we call Positional Variance Profiles (PVP); these are naturally adapted to monitor variability, require almost no parsing, and can handle intermittent and continuous control equally well.

6.1.1 Computation of PVPs

The idea of PVPs is to monitor how the positional variance of the limb extremity varies over time during a fixed observation window, say 2 seconds.

In PVPs, starting times are determined precisely, but stopping times need not be. As shown in Fig. 2.5, there remain long quasi-stationary periods at the end of the movement, when the limb stabilizes and the final position is maintained, which usually makes determining exact stopping times difficult.

We consider these stationary period for as long as they hold, if the movement drastically changes before the 2 second period e.g., because the participant moved his hand back to the starting position, we artificially extend the movement so that it lasts as long as the observation window. An example is given in Fig. 6.1-a. The participant reached the line in about 0.6 seconds, and kept its position for about an extra 0.3 seconds, after which he lifted his hand up (blue curve). The movement is then extended to fill the observation window (orange curve).

The trajectories obtained this way are then synchronized by using the starting time as origin. Fig. 6.1-b displays a set of synchronized and extended trajectories. Notice that after about 1 second, the trajectories are all stationary. The trajectories display a great amount of variability, which justifies the stochastic approach (see 2.1.1).

To obtain the PVP, we then simply compute the positional variance using all trajectories at each time instant, see Fig. 6.1-c. This method not only dispenses the complicated task of determining stopping times, but can also handle trajectories which do or do not display submovements, thereby also solving the difficulty raised in 2.1.3. An additional advantage is that an analysis in terms of variance does not suffer from noisy differentiations necessary to compute speed or acceleration profiles [44].

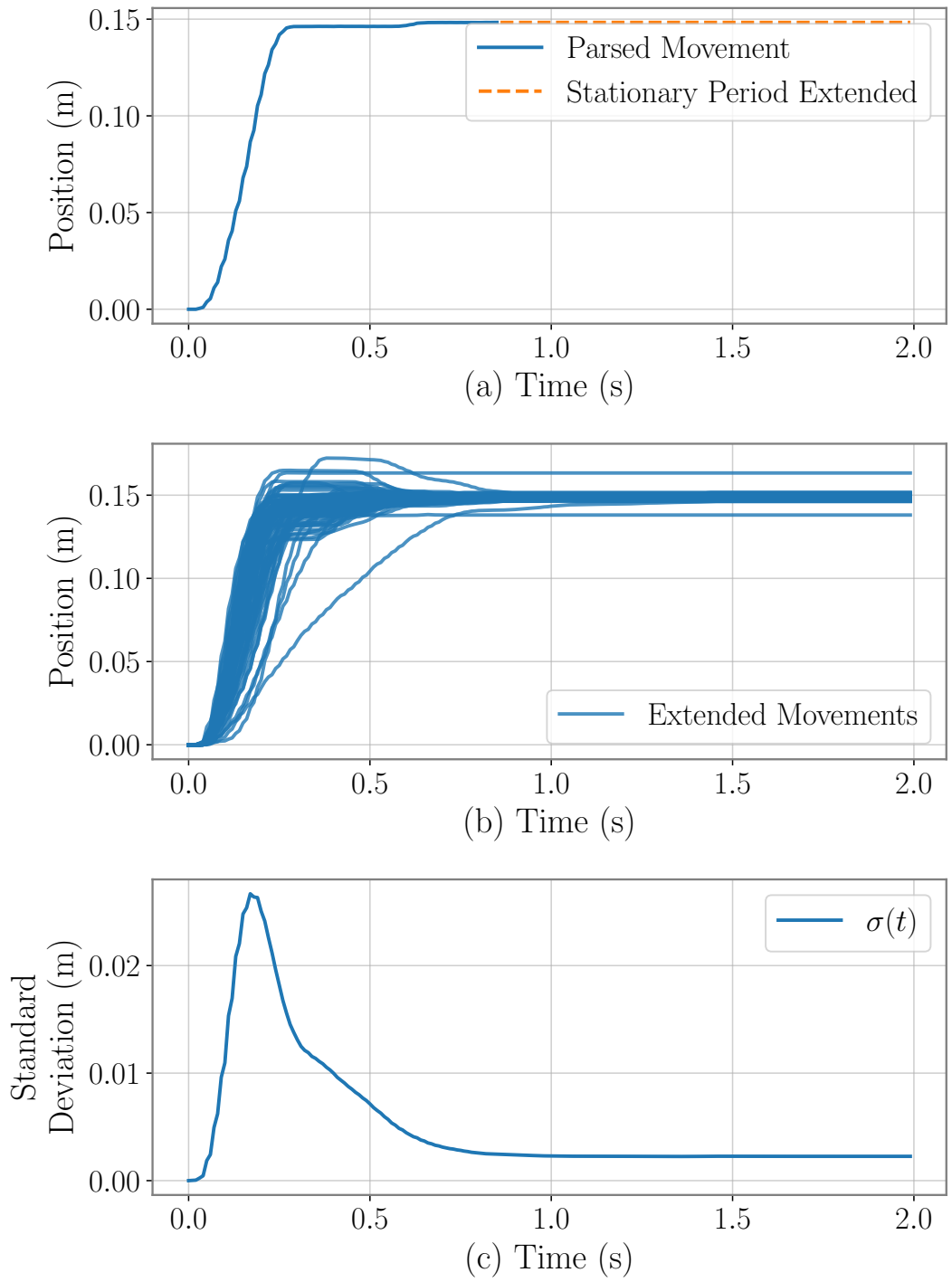

Figure 6.1: (a) One movement produced by a participant in a tapping task. The parsed movement is displayed in the blue line, while the extended portion is given in the orange dashed line. (b) A set of movements produced by the same participant performing in the same condition as (a). (c) PVP computed from (b).

6.1.2 Conjecture: Unimodality of PVP

Some previous works have studied whole trajectories, e.g., [87] evaluates the temporal evolution of the entropy of trajectories from a tapping task. Entropy profiles, as well as profiles of standard deviation of position were reported *unimodal* (an increasing phase, followed by a decreasing phase); this is consistent with the PVP of Fig. 6.1-c. Other studies [84, 149] have represented positional variance at specific kinematic markers (peak acceleration, peak velocity, peak deceleration, movement time). Extrapolation of data again suggests unimodal PVPs.

However, the variability of trajectories produced by elbow flexion are actually *bimodal* [69, 23]. One reason for the discrepancy between elbow flexion studies and the other studies mentioned is that elbow flexion allows only one degree of freedom, whereas the other studies allowed several joints to be mobilized. The following simple reasoning tends to justify unimodality for tapping movements:

1. all movements starting from the same position, initial positional variance is null (Fig. 6.1-b at 0 s);

2. in the early stages of the movement, positional variance necessarily increases [129, 23] (Fig. 6.1-c from 0 to about 0.2 s);

3. eventually, the target is reached and the movement ends. The fact that humans are capable of reaching a target reliably, whatever its size, implies that the positional variance can be reduced at will [160] (Fig. 6.1-c from about 0.2 s to 2 s), and in studies of targeted arm movements, reduction in variability as the target is approached have been observed consistently [141, 140, 23].

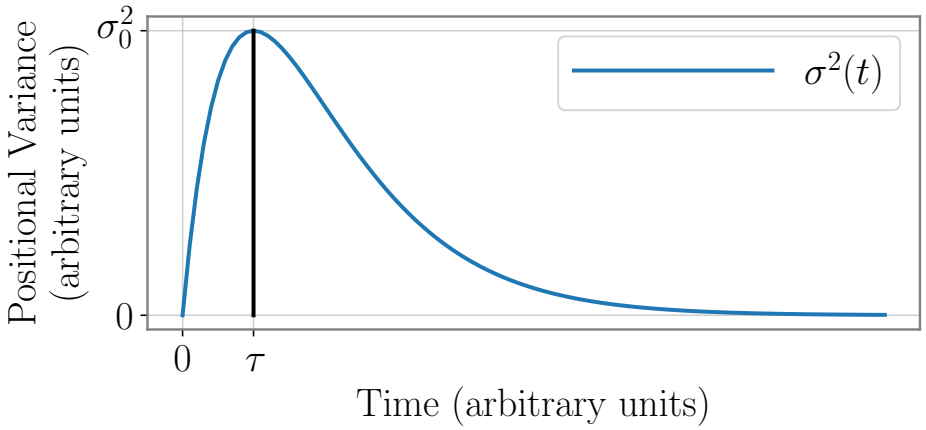

Figure 6.2: Ideal two-phase positional variance profile. The transition between the two phases occurs at $(\tau; \sigma_0^2)$.

We therefore expect *unimodal* PVPs as illustrated in Fig. 6.2: A first phase, for time

$\in [0, \tau]$, where variance *increases* from 0 to σ_0^2 is followed by a second phase, where variance *decreases* to arbitrarily low values.

PVPs will be evaluated in the next part; for now we simply assume the unimodality of the PVPs. As the second phase of decreasing phase is much longer than the first one, the second phase can be expected to dominate overall movement time, hence studying only this phase will be a good enough approximation for movement time. In Chapter 9, we not only show this assumption to be acceptable, but more importantly that the first phase has approximately a constant duration.

6.2 A Model for the Variance-Decreasing Phase

In the following, we propose an information-theoretic model for the second phase, where variance decreases.

6.2.1 Information-Theoretic Model Description

While most previous models aim at predicting an average trajectory without attention for its variability, our goal here is to predict how the variance of a set of trajectories evolves over time. Asymptotically (for large enough datasets) the position of the limb extremity ('limb' in short) is essentially random. We assume that, at the end of the first phase, it can be well approximated by a Gaussian distribution, with some standard deviation σ_0 (see Fig. 6.2).

This Gaussian assumption is often used [129, 102], as it is an idealization of actual inputs that is manually tractable and likely widely applicable [31]. The assumption depends on the conditions under which movements are produced [71]; with extreme emphasis of speed or accuracy usually a skew appears in the estimated distributions. We verified that in the second phase, position distributions were approximately symmetric. In [87], the empirical entropy of the trajectory was compared to that of the Gaussian distribution; the difference was never more than 0.3 bits throughout the entire trajectory.

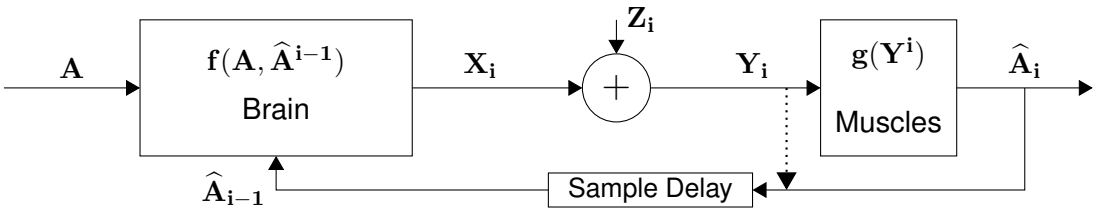

Figure 6.3: Information-theoretic model for the aiming task, with A the initial distance to the target at the end of the first phase; X_i the new distance signal created by the brain at $t = \tau + iT$; Y_i the noisy version of X_i, perturbed by Gaussian noise Z_i such that $Y_i = X_i + Z_i$; $\widehat{A^i}$ the distance that is actually covered by the limb at $t = \tau + iT$; and T the time between two iterations.

Thanks to feedback information (see 2.1.2), the limb position is known at the brain level. Due to eye-hand coordination and fast eye dynamics, the eye is usually pointing towards the target long before the end of the movement [33]. Hence the position of the target is also known at the brain level. The distance from limb to target can thus be easily evaluated by the brain; in fact it can be readily estimated by the eye if the limb is close enough to the target. Distance estimated from visual extra information about the limb and compared with visual information about the position of the target is key [115] and aleviates the need to planifiy precisely the effects of e.g. coriolis and gravitational forces [140].

Let A be the distance from limb to target; from the above assumptions, A is modeled as a Gaussian random variable with standard deviation σ_0. Movements may equally well undershoot or overshoot the target [102] when users are instructed to perform fast movements. We assume zero mean: $A \sim \mathcal{N}(0, \sigma_0^2)$. To complete the movement, the brain has to send A to the limb; therefore the role of the second phase is essentially to send a real number from a Gaussian source (A) from the brain (emitter) to the limb extremity (destination). Consider the following scheme:

- From A the brain outputs an amplitude X_1 to be sent to the limb:

$$X_1 = f(A), \tag{6.1}$$

  where f is some function performed by the brain.

- To account for the variability of the human motor system 2.1.1, we consider a noisy transmission from brain to limb, where X_1 gets perturbed by additive white Gaussian noise (this constitutes a so-called AWGN channel). The output of the channel Y_1 is given by

$$Y_1 = X_1 + Z_1 \quad \text{where } Z_1 \sim \mathcal{N}(0, N), \tag{6.2}$$

  where N is the power of the noise.

- Based on the channel output Y_1, the distance $\widehat{A}_1$ is actually covered, which is the result of some function g applied by the motor organs to the received Y_1.

- $\widehat{A}_1$ is returned to the brain *via* ideal (noiseless) feedback where it is compared to A. From such a comparison a new amplitude X_2 is produced by the brain.

The scheme then progresses iteratively for $i = 1, 2, \ldots$. We assume that each step i, from the creation of X_i to the reception of $\widehat{A}_i$ takes T seconds. If each such step becomes infinitesimally small, the whole process becomes an intermittent iterative correction model, that is continuous at the limit.

At iteration i, the scheme is described by following equations (see Fig. 6.3):

1. The brain (the 'encoder') produces X_i from A and all received feedback information $\widehat{A}^{i-1}$: $X_i = f(A, \widehat{A}^{i-1})$.

2. The motor organs (decoder) receive Y_i contaminated by Gaussian noise: $Y_i = X_i + Z_i$.
3. The covered distance $\widehat{A}_i$ is a function of all previous received amplitudes: $\widehat{A}_i = g(Y^i)$.

At this stage, f and g are still undetermined.

In Shannon's communication-theoretic terms, the aiming task in the second phase can thus be seen as the transmission of a real value from a "source" (distance from target at the end of the first phase) to a "destination" (limb extremity) over a noisy Gaussian channel with noiseless feedback. In human-centered terms, the second phase is the one which deals specifically with aiming—to make sure that the limb *reliably* reaches the target, once most of the distance has been covered. This second phase likely corresponds to the submovements of the stochastic optimized submovement model [102], and to the second component in two-component models [167, 34, 33].

6.2.2 Bounds on Transmitted Information

The goal of this subsection is two-fold. First, we would like to compute the maximal rate of transmitted information. Second, we would like to determine the the unknown functions f and g of the *optimal* scheme i.e., the one which attains this maximum. To do so, we now leverage information-theoretic definitions.

- $P_i = \mathbb{E}[X_i^2]$, where $\mathbb{E}$ is the mathematical expectation, is the input's average power.
- The quadratic *distortion* $D_n = \mathbb{E}[(A - \widehat{A}_n)^2]$ is the mean-squared error of the estimation of A by $\widehat{A}_n$ after n iterations (channel uses); this essentially corresponds to the empirical variance.
- $I(A, \widehat{A}_n)$ is Shannon's mutual information [18] between A and $\widehat{A}_n$.
- Shannon's capacity C [132, 18] of the AWGN Channel under power constraint $P_i \leq P$ and noise power N is

$$C = \frac{1}{2}\log_2(1 + P/N), \tag{6.3}$$

  expressed in bits per channel use.

The first result comes as a bounds on mutual (transmitted) information.

Theorem 6.2.1. *Consider the transmission scheme of Fig. 6.3 with an AWGN channel of capacity C and noiseless feedback. For a zero-mean Gaussian source A with variance σ_0^2, we have*

$$\frac{1}{2}\log\frac{\sigma_0^2}{D_n} \leq I(A, \widehat{A}_n) \leq nC. \tag{6.4}$$

The proof is given in the Appendix. The Theorem expresses that enough information should be transmitted from the brain to the limb to reduce the positional variability from the initial variance (σ_0^2) to the variance at the end of the movement (D_n). However, it also expresses that the transmitted information can never exceed nC, where n is the number of iterations. Since the rate per iteration can never exceed C, being more accurate requires sending larger amounts of information, which in turn requires more iterations of the scheme. In the end this leads to higher movement time, hence this inequality already poses the premises of the speed-accuracy tradeoff.

This inequality can be easily adapted to the case where A is arbitrarily distributed. In that case, σ_0^2 in (6.4) should be replaced by

$$\widetilde{\sigma}_{\mathsf{A}}^2 = 2^{2H(A)} \tag{6.5}$$

where $H(A)$ is the entropy of A and $\widetilde{\sigma}_{\mathsf{A}}^2$ is the so-called entropy power of A.

6.2.3 Achieving capacity

For a given channel and number n of channel uses, maximizing accuracy is equivalent to minimizing D_n. Similarly, for a given accuracy (distortion), minimizing time is equivalent to minimizing n. Optimal aiming, which consists of achieving the best possible accuracy in the least amount of time is thus achieved when equality holds in (6.4):

$$\frac{1}{2}\log\frac{\sigma_0^2}{D_n} \;=\; I(A,\widehat{A}_n) \quad = \quad nC. \tag{6.6}$$

The goal of this subsection is to find the scheme that achieves optimality (6.6). Before heading on, it is important at this point to observe the following. A well-known difficulty with information-theoretic models for human performance is that usually information is not directly measurable; hence proposed models are usually not falsifiable, and thereby not admissible according to critical rationalism [119]. But notice from (6.6) that on one hand, mutual information $I(A,\widehat{A}_n)$ can be estimated directly from the ratio of two measurable variances (σ_0^2 the initial variance of the set of trajectories, and D_n the variance of the set of trajectories after n iterations), and on the other hand that mutual information can actually be removed from the equality in (6.6); the usual pitfall is thus evaded.

In all what follows we use the list notation $\ell^i = (\ell_1, \ldots, \ell_i)$.

Lemma 6.2.1. *Optimal aiming can be achieved if, and only if, we have the following conditions:*

1. *all considered random variables A,$\widehat{A}^i$, $A-\widehat{A}^i$, X^i, Y^i, Z^i are Gaussian;*
2. *input powers $P_i = \mathbb{E}[X_i^2]$ are equal (to, say, P);*
3. *endpoints $\widehat{A}^i$ are mutually independent;*
4. *channel outputs Y^i and errors $A-\widehat{A}^i$ are independent;*

5. $\widehat{A}_i = g(Y^i)$ *is a sufficient statistic of* Y^i *for* A.

The proof is given in the Appendix. The fact that all random variables are Gaussian in the optimal scheme considerably simplifies operations, as independence between Gaussian variables is equivalent to decorrelation. It is important to remind the definitions of independency and decorrelation at this point. If X and Y are independent, then by definition $f_{X|Y}(x|y) = f_X(x)$, while if X and Y are decorrelated, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. If additionally either one of X or Y is a zero mean random variable, then $\mathbb{E}[XY] = 0$. It is easy to show that independence implies decorrelation as

$$\mathbb{E}[XY] = \int f_{XY}(xy)xy\ dxdy = \int f_{X|Y}(x|y)f_Y(y)xy\ dxdy. \tag{6.7}$$

Using the definition of independence, this simplifies to

$$\mathbb{E}[XY] = \int f_X(x)f_Y(y)xy\ dxdy = \mathbb{E}[X]\mathbb{E}[Y]. \tag{6.8}$$

However, the converse is not true as decorrelation does not necessarily imply independence. A well-known example is to consider $X = \pm 1$ with equal probability and $Y = X^2$. As the definition of decorrelated variables is easier to use than independent variables, Gaussian variables considerably simplify our analysis.

A further advantage of the Gaussian setting is that the minimum mean-squared error (MMSE) estimator —which minimizes D_n— reduces to a linear function in the Gaussian setting. This will be useful below, and ultimately leads both f and g to be linear functions. This in turn is not surprising; linear functions are known to preserve normality, which is a requisite for item 1 in Lemma 6.2.1.

If the power of the input P_i is constant during the various iterations while the power of the error signal diminishes gradually, then f has to incorporate a scaling mechanism; we will find this indeed to be true.

Finally, since $g(Y^i)$ is a sufficient statistic of Y_i for A, it does not matter if the feedback comes from the endpoints $\widehat{A}^i$ or from the outputs of the channel Y^i (see dotted arrow in Fig. 6.3). In this way our model can also account for feedback information prior to the motor organs (e.g. kinesthetic feedback).

By working out the conditions of Lemma 1, we can derive the structure of the optimal scheme, i.e. find the expression of f and g. We first obtain g by using the well-known *orthogonality principle*.

Orthogonality Principle: If the parameter x is to be estimated from the observed data y by the unbiased linear estimator $\hat{x}(y)$, the orthogonality principle states that the following are equivalent:

$$\bullet\ \hat{x}(y) = \mathbb{E}[x|y] = \mathbb{E}[xy]^t\mathbb{E}[yy^t]^{-1}y \text{ is the MMSE estimator;} \tag{6.9}$$

$$\bullet\ \mathbb{E}\left[(x - \hat{x}(y))y^t\right] = 0 \tag{6.10}$$

The term $(x - \hat{x}(y))$ represents the error in estimating x by $\hat{x}$. In our model, we estimate the parameter A from the observations Y^i, with $\widehat{A}_i$, hence the following result.

Proposition 6.2.1. *For the optimal transmission scheme, $g(Y^i)$ is the MMSE estimator: $g(Y^i) = \mathbb{E}\left[A|Y^i\right]$.*

Proof. From condition 4), $\mathbb{E}[(A - g(Y^i))Y^i] = 0$. The result follows immediately from the orthogonality principle. □

The optimal scheme thus yields an endpoint $\widehat{A}_i = g(Y^i)$ obtained as the best least-squares estimation of A from all the previous observations of channel outputs $Y^i = (Y_1, \ldots, Y_i)$. This optimal behavior is similar to studies on signal detection and stimulus discrimination [111, 61], where performance of subjects also approximates the behavior of an ideal observer who must deal with internal noise. Actually computing the MMSE in the Gaussian case is a matter of scaling, which is a function that e.g. muscles can perform easily [127].

We now turn to f.

Proposition 6.2.2. *For the optimal transmission scheme,*

$$X_i = f(\widehat{A}^{i-1}, A) = \alpha_i(A - \widehat{A}_{i-1}) \tag{6.11}$$

$$= \alpha_i(A - \mathbb{E}\left[A|Y^{i-1}\right]), \tag{6.12}$$

where α_i meets the power constraint $\mathbb{E}[X_i^2] = P$.

The proof is given in the Appendix. The signal sent to the channel is thus simply the difference between the initial *message* A and its current *estimate* $\widehat{A}_{i-1}$, rescaled to meet the power constraint (Condition 2 in Lemma 6.2.1). It is recognized that the distance between target distance and actual distance is a fundamental signal for movement control, as discussed in 6.2.1. The basal ganglia, and more precisely the pallidum, may serve to scale the amplitude of movements [127], hence α_i could result from this part of the brain.

The previous two propositions formally define the encoding function f and decoding function g; both are linear functions. The functions f and g are mathematically simple and biologically feasible and the distance difference $A - \widehat{A}^{i-1}$ can be readily estimated by the eye.

The next result shows that the procedure is incremental, yet optimal at each step, allowing optimal on-line control.

Proposition 6.2.3. *Let $A_i = X_i/\alpha_i$ be the unscaled version of X_i. We have*

$$\mathbb{E}[A|Y^{i-1}] = \sum_{j}^{i-1} \mathbb{E}[A|Y_j] = \sum_{j}^{i-1} \mathbb{E}[A_j|Y_j]. \tag{6.13}$$

Again the proof is given in the Appendix. The theorem shows that the "decoding" process is recursive: At each step, a "message" A_i that is independent from the

previous ones $A^{i-1} = (A_1, \ldots, A_{i-1})$ is formed, and is then estimated optimally by least-square minimization.

The incremental aspect of the procedure is a strong feature for the model. In most models where trajectory formation is based on the minimization of a cost function lies a problem of causality: How can the brain conceive of a trajectory that minimizes e.g. jerk [43] throughout movement, while also adapting the trajectory with feedback information ? To be more explicit, minimizing jerk involves computing a path integral, from start to finish of the movement. Supposedly the motor planning system then selects the trajectory that minimizes the jerk, but that leaves out any possible correction through feedback. Hence models that minimize some cost over the whole movement can only predict data, but never explain movement generation.

Because in our model the path is incrementally and globally optimal, it suffices for the participant to do as best as he can at each iteration and this will lead to the best possible scheme for the whole movement.

Finally, we verify that the obtained distortion is indeed the one obtained in the optimal scheme (6.6):

Theorem 6.2.2. *The quadratic distortion $D_i = \mathbb{E}[(A - \widehat{A}_i)^2]$ decreases exponentially in i:*

$$D_i = \frac{\sigma_0^2}{(1 + P/N)^i}. \tag{6.14}$$

With this scheme, it is immediately checked that capacity C is exactly achieved and the distortion decreases geometrically (divided by $1 + P/N$) at each iteration. The scheme successfully matches the transmission of one real value A using feedback with n independent channel uses (transmissions).

The above equations were already obtained in an information-theoretic context by Gallager and Nakiboğlu [47] who discussed an older scheme by Elias [29]. The constructive approach of the Elias scheme given here, as well as its application to model human aimed movements is novel.

The model will be validated in the next part, but at this point it is useful to showcase an example. Fig. 6.4 shows PVPs in logarithmic y-scale. As the variance is expected to decrease exponentially in time during the second phase, it should appear as a linear decreasing portion when the y-axis is expressed in a log scale. The decrease from the point of maximum variability (τ, σ_0) is indeed linear; it is much longer for conditions emphasizing accuracy. The final stationary period, where movements are finished corresponds to the horizontal part. The transition between the first and second phase, and between the second and stationary phase is spread out — or rounded off; this should be accounted for by curve fitting.

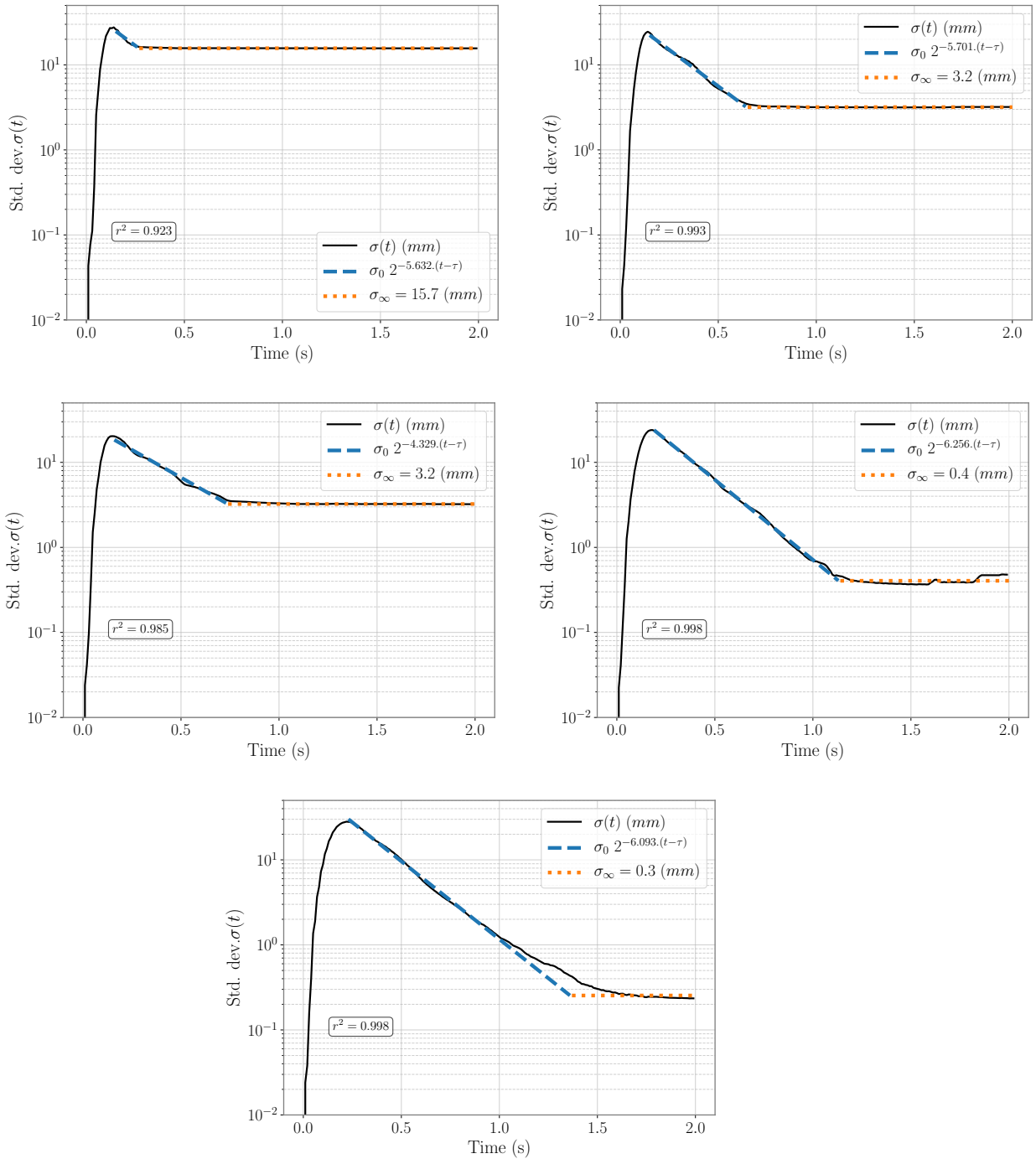

Figure 6.4: PVPs in a log-y scale. Data from one participant of the G-dataset, for all conditions. Top left is the maximum speed emphasis; bottom right is maximum accuracy emphasis (Instructions to be more precise from left to right and from top to bottom). The blue fitted line is the theoretical linear prediction from FITTS 2. The orange fitted line is the stationary phase, where the movement is stopped.

6.3 Exponential Decrease of Variance: Difficulty, Throughput and Fitts' law

The scheme presented in the previous section uses a discrete-time formalism, where n in (6.6) is the number of the iterations of the scheme. If the iteration time is constant and equal to T, total time t is given by $t = nT$.

When considering many trajectories, it is likely that any time interval, however small will contain at least one new correction, providing a smooth variance profile. We can thus assume an infinitely small looping time δT so that the number of iterations $n = t/\delta T$ tends to infinity. This does not entail that each trajectory is obtained as the result of a continuous on-line control, but rather that the *entire set* of trajectories can be described by a model where variance decreases smoothly.

6.3.1 Local Exponential Decrease of Variance

We can rewrite (6.6) in terms of the second phase's duration t

$$t = \frac{\delta T}{1/2 \log(1 + P/N)} \log_2 \frac{\sigma_0^2}{D_n} \tag{6.15}$$

$$= \frac{\delta T}{C} \frac{1}{2} \log_2 \frac{\sigma_0^2}{D_n} \tag{6.16}$$

$$= \frac{1}{C'} i_d, \tag{6.17}$$

where we define i_d as the *local* index of difficulty for the second phase:

$$i_d = \frac{1}{2} \log_2 \frac{\sigma_0^2}{D_n} \tag{6.18}$$

and

$$C' = C/\delta T \tag{6.19}$$

the capacity in *bit per second*. This is very similar to Fitts' law (1) without intercept. For practical purposes, we have $D_n = \sigma^2(t)$; rearranging terms:

$$i_d = \log_2 \frac{\sigma_0}{\sigma} = C'\, t, \tag{6.20}$$

This is a *local* formulation for Fitts' law that holds between any two dates in the second phase: For arbitrary small $\Delta t \geq 0$,

$$\log_2 \sigma(t + \Delta t) = \log_2 \sigma(t) - C' \Delta t. \tag{6.21}$$

Thus, the second phase for optimal movements is characterized by a linearly decreasing standard deviation of position where the slope is exactly the inverse of the capacity

of a Gaussian channel. Although PVP can be equivalently computed by considering the variance or the standard deviation of position over time, it is advantageous to consider the standard deviation whose slope gives the direct estimate of C'.

This local formulation is a real space-time formulation for the speed-accuracy trade-off that links time at any instant (t) to accuracy at any instant ($\sigma(t)$), and in this respect is a much fuller description than Fitts' law, which only predicts time at the end of a movement for one given accuracy level.

6.3.2 Local Index of Difficulty i_d and Throughput

The local index of difficulty i_d is cleanly expressed as a ratio between initial and final variance; it relates to the physical decrease of variance necessary (from σ_0^2) to achieve reliable pointing (to D_n). Thus, i_d is easily interpretable beyond the vague notion of bits.

In fact, i_d solves the problem of the null ID discussed in 3.2.3. With i_d, a set of movements of zero difficulty is one where the variance is not changed, i.e. no movement needs to be performed.

Many researchers have attempted to derive ID as the difference between an initial and a final entropy [56]. With the local, i_d can precisely be expressed as the difference between two entropies:

$$i_d = \frac{1}{2}\log 2\pi e\sigma_0^2 - \frac{1}{2}\log 2\pi e D_n, \tag{6.22}$$

where the term on the left of (6.22) is the entropy of a Gaussian variable with variance σ_0^2 and the term on the right of (6.22) is the entropy of a Gaussian variable with variance D_n. A crucial difference between ID and i_d is that the former is task defined, whereas the latter is not; as a result there is no distinction between a nominal or an effective index as in 3.3.

Finally, this local formulation also solves the issue of defining throughput 3.4. Here, the throughput is unequivocally defined as

$$C' = \frac{1}{2}\log_2(1 + P/N)/\delta t \tag{6.23}$$

which is the Shannon capacity in bit/s of the feedforward channel, with bandwidth

$$B_W = (2\delta t)^{-1}. \tag{6.24}$$

In the continuous scheme, the bandwidth is infinite, as δ is infinitesimally small. This does not imply infinite bandwidth for one movement, but for an infinite set of movements (remember that all our results are asymptotic, hence require an infinite number of traces).

6.3.3 Deriving the Classic Fitts' law

The local formulation (6.20) can be exploited theoretically in the context of Fitts' paradigm and should normally yield Fitts' law (1) to ensure consistency. It is apparent from (6.20) that some expression for σ_0 as a function of D and/or W is required to obtain Fitts' law.

For very short movements, Schmidt's model may be assumed:

$$\sigma_0 = k_1 + k_2 \mathrm{D}_\tau, \tag{6.25}$$

where σ_0 is the standard deviation at the end of the first phase [129] and k_1 and k_2 two constants. When plugged into (6.20), we get

$$\mathrm{MT} = \tau + 1/C' \times \log_2 \frac{k_1 + k_2 \mathrm{D}_\tau}{\sigma}. \tag{6.26}$$

If we assume Meyer's model [102], i.e. that the intercept in (6.25) can be neglected, we arrive at

$$\mathrm{MT} = \tau + 1/C' \log_2 \frac{k_2 \mathrm{D}_\tau}{\sigma}. \tag{6.27}$$

In 9.3.4, we show that the duration τ of the first phase is approximately constant. We also show that the distance traveled at τ, D_τ is about two thirds of the total distance D. Hence, (6.25) becomes

$$\begin{aligned} \mathrm{MT} &= \tau + 1/C' \log_2(2/3\ k_2) + 1/C' \log_2 \frac{\mathrm{D}}{\sigma} && (6.28)\\ &= \tau' + 1/C' \log_2 \mathrm{D}/\sigma, && (6.29) \end{aligned}$$

which is Mackenzie's version of Fitts' law 3.19, except for the "$+1$" inside the logarithm. This formula shows that the throughput TP, as defined in (3.22) as the inverse of the slope is exactly equal to C' as defined in (6.24), and is the expression of the capacity of a noisy Gaussian channel.

It is known that participants are able to keep a relative constant error rate ε (usually, ε is about 5%) across conditions [38, 40, 142]; for a Gaussian distribution of endpoints, we have that $\sigma \simeq [2\sqrt{2}\mathrm{erf}^{-1}(1-\varepsilon)]\mathrm{W}$, where $\mathrm{erf}^{-1}(x)$ is the inverse Gaussian Error Function (see 3.6 for an example of the computation). This leads to

$$\begin{aligned} \mathrm{MT} &= \tau' + 1/C' \log_2 \frac{\mathrm{D}}{2\sqrt{2}\mathrm{erf}^{-1}(1-\varepsilon)\mathrm{W}} && (6.30)\\ &= \tau' - 1/C' \log_2\big(2\sqrt{2}\mathrm{erf}^{-1}(1-\varepsilon)\big) + 1/C' \log_2(\mathrm{D}/\mathrm{W}), && (6.31)\\ &= \tau'' + 1/C' \log_2 \mathrm{D}/\mathrm{W} && (6.32) \end{aligned}$$

This is Crossman's version of Fitts' law (3.4) which is equivalent to Fitts' formulation (3.3) (see Proposition 3.2.1). Thus, the fact that Fitts' law conflates separable effects of target distance and width [99] is fully explained if one is ready to assume Schmidt's model.

6.3.4 Interpreting the Intercept

The intercept τ'' in (6.32) incorporates the duration of the first phase τ as well as constants of Schmidt's law and the miss rate ε. It is thus expected that the intercept is largely paradigm-dependent. The parameters of Schmidt's law are expected to vary significantly, depending on the degree of practice with the task [137]. The serial task in [109] leaves less time for movement preparation than the discrete task in [66], which leads to a longer τ, see Chapter 9. We expect this to be a generalizable result. Finally, τ'' incorporates ε, hence is also dependent on the strategy adopted by the participant, as well as the instructions given by the experimenter.

Definitions of throughput using τ'', such as $\overline{\mathrm{TP}}$ (3.23) are paradigm dependent, while definitions of throughput using C', such as TP (3.22) only account for the second phase of movement.

6.4 Discussion

Looking back to the beginning of this second part, the models for aimed movement have substantially evolved; Fig. 6.6 offers a useful recapitulation. The first model, presented in 4 was a simple source model from which various ID could be derived. The source model is equivalent to a noiseless transmission model.

In Chapter 5, we considered a noisy channel, first with uniform bounded noise, which we then generalized to any bounded noise. We also showed how pointing mistakes could be accounted for, by using $\mathrm{ID}(\varepsilon)$. The model does not offer many insights into the control of movement nor does it offer new predictions; it also fails to encompass feedback. This second model is in fact an attempt, to push a simple information-theoretic scheme as far as we can in an effort to reconcile Fitts' law with Shannon's information-theory. At the price of a few simplifying assumptions, the reconciliation is indeed possible. The model does have another merit, as it underlines the need to consider on one side average performances, and on the other hand extreme performances. This observation will be leveraged in Chapter 8.3.2.

The third model, detailed in this chapter, is the crux of the thesis. It is a biologically reasonable model constructed from a comprehensive analysis of human behavior, which assumes very little *a priori*. Most of the functions are left unknown at the start, and are determined solely by the fact we look to maximize the information transmitted from the brain to the limbs. A resulting implementation is as follows:

The A_i's, given by the difference between the target position A and current position $\widehat{A}_{i-1}$, represent the distance that remains to be covered. This distance can readily be evaluated by the eye; because of eye-hand coordination and fast eye dynamics, the eye is usually pointing towards the target, giving high fidelity spatial information to the brain [33]. The distance signal is rescaled (multiplied by α_i) and sent through the nervous system (the "channel") where it gets corrupted by additive noise Z_i; this noise encompasses the variability characteristic of human aimed movement. Between the channel output and the limb, MMSE estimation is performed, which amounts to

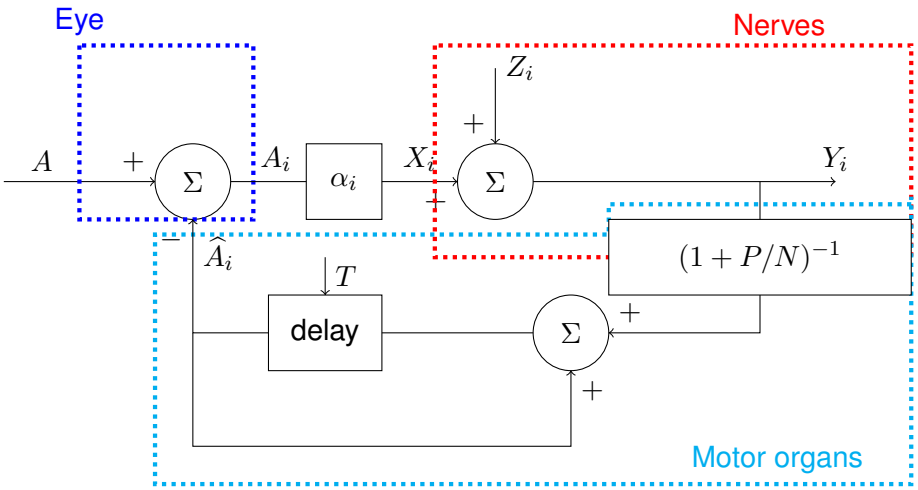

Figure 6.5: Possible implementation of how human perform the aiming task: functions and potential organ groups who can perform these functions.

multiplying the channel output Y_i by the ratio between the variance of the input signal P and the variance of the output signal $P + N$. Which organs could actually perform this function which relies on the knowledge of P and N is unknown. The position of the limb extremity is then updated by adding the newly decoded output to the previous ones. This operation is represented by the loop at the bottom of Fig. 6.5. The loop and the delay constitute a memory used to store the current position (the limb extremity remains still if no signal is sent). This model describes the variations of standard deviation over the whole trajectory which is quite an improvement over Fitts' formula which can only advise on movement endpoints.

The assumption that the source of information can be represented by a zero-mean variable holds quite well for balanced strategies [71], but fails for extreme emphasis of either speed or accuracy. Remarkably, we will see in Chapter 9 that the extreme accuracy condition actually provides the best fit with our model. Elliott et al. [35] provide a tentative explanation. They argue that is it likely that the two components (or two phases) do not necessarily occur serially, but can overlap as well. If the feedforward information i.e. the average value of A is given to f, then f can then simply subtract the known average value to consider once again the zero-mean variable.

An important feature of this model is that it is locally optimal, as the optimal solution is actually constructed at each iteration from a local optimizing problem. This contrasts with other models who work by minimizing a cost constraint over an integral path, such as the minimum-jerk models [43]. Indeed, if a mechanism chooses a path by minimizing a quantity over the whole trajectory, then how can it change its path via feedback information ? Conversely, how can a controller compute the optimal path using information that will only be available in the future once the movement has already started ? Models who minimize a global cost hence have either a problem accounting for feedback or are uncausal. Proposition 6.2.3 thus constitutes a strong point of FITTS 2.

In the next part, we look at empirical considerations for both transmission models.

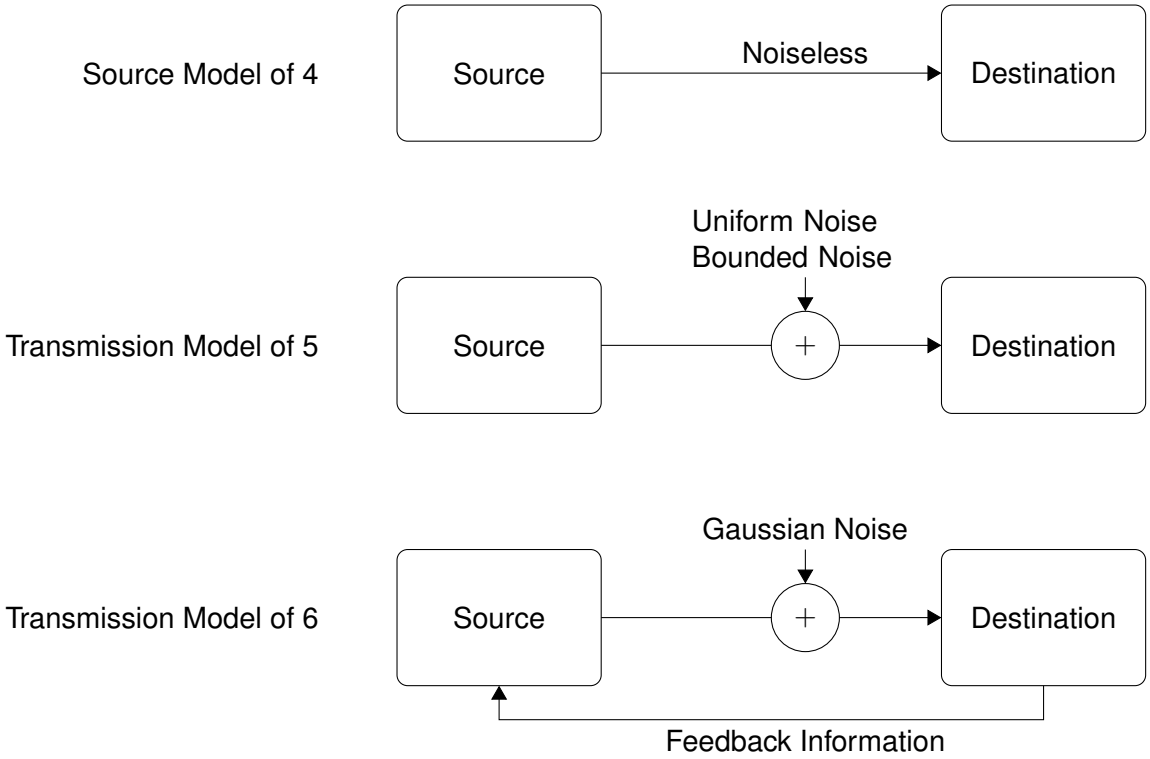

Figure 6.6: Recapitulation of the 3 models of Part II.

For the first model there are no new predictions, hence nothing to validate, but the front of performance does call for new tools and techniques, which we will describe. The second model will be thoroughly validated, and we will see that it actually induces Fitts' law at the endpoints.

6.5 Appendix

Proof of Theorem 6.2.1. The proof uses well known techniques and inequalities from information-theory [18]. For the first inequality in (6.4):

$$
\begin{aligned}
I(A;\widehat{A}_n) &= H(A) - H(A|\widehat{A}_n) && (6.33)\\
&= H(A) - H(A-\widehat{A}_n|\widehat{A}_n) && (6.34)\\
&\geq H(A) - H(A-\widehat{A}_n) && (6.35)\\
&\geq H(A) - \frac{1}{2}\log\left(2\pi e\mathbb{E}[(A-\widehat{A}_n)^2]\right) && (6.36)\\
&= \frac{1}{2}\log\frac{\sigma_0^2}{D_n} && (6.37)
\end{aligned}
$$

where we have used the definition of mutual information (6.33); the fact that conditioning reduces entropy (6.35); the fact that the Gaussian distribution maximizes entropy under power constraints (6.36); and the definition of distortion and the entropy formula for a Gaussian distribution (6.37). For the second inequality in (6.4):

$$
\begin{aligned}
I(A;\widehat{A}_n) &\leq I(A;Y^n) && (6.38)\\
&= H(Y^n) - H(Y^n|A) && (6.39)\\
&= \sum_i \left[H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1},A)\right] && (6.40)\\
&= \sum_i \left[H(Y_i|Y^{i-1}) - H(Y_i|X_i)\right] && (6.41)\\
&\leq \sum_i \left[H(Y_i) - H(Z_i)\right] && (6.42)\\
&\leq \sum_i \left[\tfrac{1}{2}\log(2\pi e(P_i+N)) - \tfrac{1}{2}\log(2\pi eN)\right] && (6.43)\\
&\leq \sum_i \frac{1}{2}\log\left(1+\frac{P_i}{N}\right) \leq nC && (6.44)
\end{aligned}
$$

where we have used the data processing inequality [18] applied to the Markov chain $A \longrightarrow Y^i \longrightarrow g(Y^i) = \widehat{A}^i$ in (6.38); the chain rule to both entropies (6.40); the feedback scheme in (6.41); the fact that conditioning reduces entropy and independence of X_i and Z_i in (6.42); that fact that the Gaussian distribution maximizes entropy in (6.43) (where P_i and N are the powers of respectively X_i and Z_i); the concavity of the logarithm function (6.44). □

Proof of Lemma 6.2.1. The proof consists of finding the conditions that turn the inequalities in the proof of Theorem 6.2.1 into equalities. Equality in Eqs. (6.35) and (6.38) directly imply condition 4. Equality in (6.36) implies that the $A - \widehat{A}^i$ are Gaussian. Equality in (6.38) implies that $H(A|Y^i) = H(A|Y^i, g(Y^i)) = H(A|g(Y^i))$, so that $Y^i \longrightarrow g(Y^i) \longrightarrow A$ should form a Markov chain, implying condition 5. Equality in (6.42) implies condition 3. Equality in (6.43) implies that the Y_i's are Gaussian. Equality in (6.44) implies condition 2 by concavity of the logarithm. Finally, X_i is Gaussian, as the result of the sum of Y_i and Z_i, both Gaussian. Similarly, $\widehat{A}_i$ is Gaussian as both A and $A - \widehat{A}_i$ are Gaussian, which finally yields condition 1. □

Proof of Theorem 6.2.2.

- On one hand, $X_i = \mathbb{E}[X_i|A, \widehat{A}^{i-1}]$ is a linear function of A and $\widehat{A}^{i-1}$, because the conditional expectation is linear for Gaussian variables.
- On the other hand, condition 3 of Lemma 6.2.1 reads that $\mathbb{E}[\widehat{A}_i\widehat{A}^{i-1}] = \mathbb{E}[g(X_i + Z_i)\widehat{A}^{i-1}] = 0$. As Z_i is independent from X_i and $\widehat{A}^{i-1}$, and using the linearity of g, we get $\mathbb{E}[g(X_i)\widehat{A}^{i-1}] = 0$.

Combining the two results, and because g is linear, we can write $g(X_i)$ as $\alpha_i(A - \tilde{f}(\widehat{A}^{i-1}))$, hence $\mathbb{E}[\alpha_i(A - \tilde{f}(\widehat{A}^{i-1}))\widehat{A}^{i-1}] = 0$. The orthogonality principle clearly appears and $\tilde{f} = \mathbb{E}\left[A|\widehat{A}^{i-1}\right]$ is the MMSE estimator. □

Proof of Theorem 6.2.3. The goal is to evaluate $\mathbb{E}[A|Y^{i-1}]$. We first use the operational formula for the conditional expectation (6.9) $\mathbb{E}[A|Y^{i-1}] = \mathbb{E}(AY^{i-1})^t\mathbb{E}[Y^{i-1}(Y^{i-1})^t]^{-1}Y^{i-1}$. Using the fact that the Y's are independent and have power $P + N$, we get:

$$\mathbb{E}[A|Y^{i-1}] = \sum_{j=1}^{i-1}(P+N)^{-1}\mathbb{E}[AY_j]\ Y_j.$$

Second, it follows from previous computations that $A - A_i = \mathbb{E}\left[A|Y^{i-1}\right]$, which is a function of the observations Y^{i-1} and therefore independent of Y_i. This leads to $A - A_i$ being independent of Y^i and the following equality: $\mathbb{E}[AY_i] = \mathbb{E}[A_iY_i]$.

Combining both results, we thus get

$$\mathbb{E}[A|Y^{i-1}] = \sum_{j=1}^{i-1}\mathbb{E}[A|Y_j] = \sum_{j=1}^{i-1}\mathbb{E}[A_j|Y_j]. \quad \square$$

Proof of Theorem 6.2.2. First notice that we can write D_i as $\mathbb{E}[(A_i - \mathbb{E}[A_i|Y_i])^2]$ as $A - \widehat{A}^i = A - \widehat{A}^{i-1} - \mathbb{E}[A_i|Y_j] = A_i - [A_i|Y_j]$. Next, we have that

$$D_i = \mathbb{E}(A_i^2) + \mathbb{E}[\mathbb{E}^2(A_i|Y_i)] - 2\mathbb{E}(A_i\mathbb{E}(A_i|Y_i))$$

From the proof above,

$$\mathbb{E}[A_i|Y_i] = \mathbb{E}[A_iY_i](P+N)^{-1}Y_i = \frac{1}{\alpha_i}\frac{P}{P+N}Y_i.$$

With some calculus and using $\mathbb{E}(A_i^2) = D_{i-1} = P/\alpha_i^2$, we get that

$$D_i = \frac{D_{i-1}}{1 + P/N}.$$

The proof is finished by applying this equation recursively. □

Part III

Leveraging and Validating the Transmission Schemes

Chapter 7

Datasets

This chapter details the datasets used throughout the thesis. Since many pointing studies have been conducted over the years, a fair share of data is available. I wish to take the time here to thank all the authors who have made access to their data possible and wish to encourage authors to routinely upload their data for everyone. Some datasets could have been spared, as there is some redundancy in the experimental procedures. However, I have tried to diversify the datasets used, to show a more general picture of pointing. I also hope this small sample of datasets can serve as a starting point for a greater pooling of pointing datasets that any researcher could easily access.

The models developed have been so in a 1-D setting, mostly for analytical convenience. Some datasets have effectively been conducted on a 1-D task (such as Fitts' original study [38]), e.g. the PD-dataset or the G-dataset. However, pointing takes place realistically in a 2-D setting. FITTS 1 and FITTS 2 could be extended to the 2-D case, but this remains future work at this point. As 2-D pointing studies suggest, various formulations for ID can be assumed without compromising Fitts' relationship severely [97], and the observations in the 1-D setting usually extend to the 2-D setting [142]. The empirical validation of FITTS 2 is conducted on 1-D datasets only, but for illustrations and perspectives, the freedom was taken to explore 2-D datasets. For FITTS 1, the definition of ID was simply replaced by a 2-D version, while for FITTS 2, the signal considered was the distance rather than the amplitude (both are the euclidean distance of the signal).

7.1 The Goldberg-Faridani-Alterovitz (GFA) Dataset

The full description of this dataset is given in [52]. Two studies were performed; the first is a controlled, in-lab study and the second an uncontrolled, web-based study. Only the data from the so-called heterogeneous task was used in this work, where the conditions of size and distance of the target vary at each trial. A 2-D Fitts task with circular targets was presented to the participants, and time was measured between the instant when the target was presented and the instant when the participant *successfully* clicked on

the target. The controlled study was run with 46 participants, each performing 10 blocks of 25 trials. The web-based study reports data from more than 2000 participants each performing a single block of 25 trials. This dataset is available on-line as detailed in the original publication.

7.2 The Guiard-Olafsdottir-Perrault Dataset (G-Dataset)

This dataset is essentially the one described in [66], which uses a discrete version of the tapping task [40]. Pointing was done towards a fixed line located at $\mathrm{D} = 150$ mm, perpendicular to the direction of movement. The participants were instructed to point towards the line while varying their strategy through five conditions: from full speed emphasis (1) to full accuracy emphasis (5). Pointing data was acquired on a high-resolution graphical tablet. This dataset was later extended using the same apparatus and software with several other participants, including two young dyspraxic participants (young lemur and young wolf, see Conclusion). In both cases, the principal investigator was Guiard, hence the name of the dataset. This dataset was gracefully shared by the authors.

7.3 The Jude-Guiness-Poor (JGP) Dataset

Details of this dataset can be found in [81]. The task used is a so-called 2-D ring-of-circles task [1], with 4 levels of D and 3 levels of W. Data comes from 15 different participants over 3 days. This dataset is only used for illustrative purposes at the start of this work and the rest of the details may therefore be skipped. The dataset is available on-line.

7.4 The Chapuis-Blanch-Beaudouin-Lafon (CBB) Wild Dataset

This dataset resulted from a large scale field study of pointing [14]. All cursor movements were unobtrusively logged over several months for 24 participants. Over 2 million 2-D pointing movements were analyzed and segmented. For each movement, a variety of information is available regarding the pointing environment, the target and the windows. Detailed information regarding the segmentation of the movement, which is naturally complex due to the nature of a field study, is given in [14]. This dataset was gracefully shared by the authors.

7.5 The PD-Dataset (Müller-Oulasvirta-Murray-Smith)

The Pointing Dynamics Dataset (here-after PD-dataset) [109] used a reciprocal (serial) version of the 1-D tapping task, where the cursor is moved back and forth between

two targets of size W separated by a distance D. The two factors were D and $\mathrm{ID} = \log_2(1 + \mathrm{D}/\mathrm{W})$; factor D had two levels D = 0.212, 0.353 (m) and ID has four levels $\mathrm{ID} = \{2, 4, 6, 8\}$ (bit), and were fully crossed. The dataset releases full trajectories, essential to validate FITTS 2.

7.6 The Blanch-Ortega (BO) Dataset

This dataset originates from a study to compare different pointing techniques in a modified version of the Fitts task where so-called distractors are included [**?**]. Indeed, some techniques can take advantage of the fact that there the display space is usually a sparse representation. Whereas in the motor space every area matters equally, in a display there are only so many icons, and usually a screen is not covered in targets. In the case of Fitts' paradigm, there is a single target, hence techniques that are not target agnostic can evidently take advantage of this. In the extreme scenario, it can be argued that with a single target to hit there is actually no uncertainty and a technique can thus instantly place the cursor on top of this target. To account for this reality, Blanch and Ortega added dummy targets, so-called distractors, and evaluated different pointing techniques. Full trajectories are released in the dataset for 5 different techniques: 1) Flat pointing (no CD gain) 2) RakeCursor [7], 3) Semantic pointing [6], 4) BubbleCursor [63] 5) DynaSpot [15]. The dataset was gracefully shared by the authors.

Chapter 8

FITTS 1: Leveraging the Front of Performance

The concept of the front of performance was established in Section 5.4. Ideally, the front of performance should have the following properties:

1. It specifies the parameters of a straight line,
2. It takes into account only the points corresponding to the best performance, through the exclusive consideration of the fronts of the performance distributions. The technique does not just discard "slow" outliers—movements of unusually long duration—it ignores all data points, but the very best.
3. It also eliminates "fast" outliers, i.e. movements of abnormally short duration.

8.1 Parametric Estimation

We build the model using each of the items listed above.

1. Fitts' law, in its front version (see Section 5.4) as derived from Shannon's capacity in Chapters 5 and 6 is expressed as

$$y = \beta_0 + \beta_1 \, x, \tag{8.1}$$

   where x is ID, y is MT, and β_0 and β_1 the intercept and slope. We carefully change notations to discriminate the front version of Fitts' law from the usual version given in (1.17).

2. In the usual linear regression model (see Chapter 3), a Gaussian noise is added to the theoretical model:

$$y = \beta_0 + \beta_1 \, x + \varepsilon, \tag{8.2}$$

with

$$\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2). \tag{8.3}$$

This assumes symmetrical errors, i.e. it is as likely that the user over or underperforms. However, a user can only do worse than (8.1), hence user data should be accounted for by a definite-positive random variable, say A:

$$y = \beta_0 + \beta_1\, x + A. \tag{8.4}$$

3. Finally, to account for deviations from (8.4), we add a Gaussian error term, similarly to the linear regression model, yielding

$$y = \beta_0 + \beta_1\, x + A + \varepsilon. \tag{8.5}$$

The deviation ε is taken to be a centered Gaussian random variable with variance σ^2. To make things easier, we can incorporate the deterministic relation (8.1) inside ε

$$\varepsilon \sim \mathcal{N}(\beta_0 + \beta_1\, x, \sigma^2), \tag{8.6}$$

leaving the following model for movement time:

$$y = A + \varepsilon. \tag{8.7}$$

The choice of A is not an easy one: to our knowledge not a lot is known about the scatter plot coming from a Fitts plot, seeing as usually the data gets averaged before any analysis is performed (see 3.1.2). A key element for A is that its standard deviation should increase with levels of x:

1. It is established that higher levels of ID lead to higher average values of MT in controlled and field studies [14]. This is in fact nothing more than the traditional Fitts' law (1.17).

2. In most psychophysics observations, the standard deviation of a quantity increases with its average value. When this scaling is linear, it is sometimes abusively[1] referred to as Weber's law.

3. It is well known that for empirical reaction times distributions, the mean is linearly related to the standard deviation [154]:

$$\sigma(\mathsf{RT}) = k\mu(\mathsf{RT}). \tag{8.8}$$

It is expected that this relation also holds for many psychological phenomenon, including movement time in a Fitts task.

[1] Weber's law, Weber's fraction or the Weber-Fechner law typically relates to perception, and relates the smallest change in stimuli that can be perceived to that stimulus' intensity — the law states that this relation is a proportional one. Note that Fitts [38] uses the term Weber fraction to qualify $\mathrm{W}/2\mathrm{D}$.

4. A recent study [81] notes that the variance of MT increases with levels of ID.

Hence, we consider a random variable A such that its standard deviation $\sigma(A)$ is an increasing function of its average $\mu(A)$. As this average is given by Fitts' law, we have that $\sigma(A)$ increases with levels of x.

Hence, A should be a positive-definite continuous distribution whose standard deviation increases with ID. This leaves several possibilities. Jude et al. [81] posit that the distribution of MT for a given ID follows a gamma distribution, albeit without explaining why. Chapuis et al. [14] fitted a type-II Generalized Extreme Value (GEV) distribution (so-called Fréchet distribution) to MT for a given level of ID, also without explanation. These two distributions have in common that they are both positively skewed.

8.1.1 Exponential Distribution

The gamma $\Gamma(k, \theta)$ law has following pdf:

$$p_{\Gamma|k,\theta}(y) = \Gamma(k)^{-1}\theta^{-k}y^{k-1}\exp(-y/\theta), \tag{8.9}$$

where $\Gamma(k)$ is the well known gamma function, which extends the definition of the factorial to the real numbers[2]. The parameter k, known as the shape, determines the rate at which the tail decreases, while the scale θ is proportional to the mean for a given k.

A distribution that is analytically simpler emerges when considering the special case $k = 1$, as $\Gamma(1, \theta)$ is simply the exponential distribution $\mathrm{Exp}(\theta)$ with pdf

$$p(y) = \theta^{-1}\exp(-y/\theta). \tag{8.10}$$

A singular fact of the exponential law is that its average value is equal to its standard deviation; namely, if $A \sim \mathrm{Exp}_\theta$ then

$$\sigma(A) = \mu(A) = \theta. \tag{8.11}$$

Hence, the distribution has the linear relationship between standard deviation and mean typical of RT distributions, albeit there is no degree of freedom to adjust the k in (8.8). However, the relative simplicity of the exponential distribution compared to the gamma or GEV distributions is appealing.

A second interesting reason to consider the exponential distribution comes once again from the maximum entropy principle; already used in Chapter 5. A is a positive-definite random variable, with infinite support, and whose mean is given by Fitts' law. The maximal entropy distribution for a random variable with a given mean is the exponential distribution [18]. Hence, only from the knowledge that Fitts' law holds for average movement time, the safest one can do is assume the exponential distribution for A.

[2] We have that the gamma function evaluated at n equals the factorial decremented by 1: $\forall\, k \in \mathbf{N}; k \geq 1, \quad \Gamma(k) = (k-1)!$.

Finally, as the average value of A is expected to increase with levels of x (ID), we write the θ parameter as x/k. Hence, the overall model is as follows:

$$y = A + \varepsilon, \tag{8.12}$$

$$A \sim \mathrm{Exp}_{x/k}, \qquad p_A(a) = k/x \exp(-ak/x) \tag{8.13}$$

$$\varepsilon \sim \mathcal{N}(\beta_0 + \beta_1\, x, \sigma^2), \qquad p_\varepsilon(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\Big(\frac{\epsilon - \beta_0 - \beta_1 x}{2\sigma^2}\Big) \tag{8.14}$$

This model has 4 free parameters $\Theta = (\beta_0, \beta_1, k, \sigma^2)$, and should be fit directly on the 2-D scatter plot. β_0 and β_1 describe the front of performance, while k describes the spread of the data. A higher value of k results in a smaller variance, which indicates that the participant is able to remain often in the neighborhood of its best performance. A low value of k means that the participant is often far away from his best performance. Hence, k can be interpreted as a measure of the regularity of a user. A good fit would leave σ^2 small compared with x/k, which indicates that the exponential law adequately models most of the scatter plot as the Gaussian deviation should just account for the few fast outliers. Note that the worst case i.e. the limiting case where the Gaussian deviation accounts for all of the data, simply amounts to computing the linear regression. A simple indicator can be given to quantify this; let

$$\alpha = 1 - \frac{\sigma^2}{\sigma_\varepsilon^2}, \tag{8.15}$$

where σ^2 is defined as in 8.14, and σ_ε^2 as in 8.3. In the ideal case where there are no deviations $\alpha = 1$, while in the limiting case where there is no exponential, $\alpha = 0$. Note that α is not a goodness of fit indicator in the sense of r^2; it is possible to imagine a case where the exponential fit is poor while simultaneously having a high α. A high α value indicates a front of performance that is rarely violated.

The parameter vector Θ can be estimated e.g. through Maximum Likelihood Estimation (MLE). Note that we not only find β_0 and β_1 which specify the front, but also k which quantifies the spread of the pointing data, as well as σ^2, needed to compute α. Basic results of MLE are recalled in the Appendix.

8.1.2 Exgauss Function

The sum of an exponential and a Gaussian function is known as an exGauss or exponentially modified Gaussian distribution, and is obtained by convolution between the pdf's of A and ε, see Appendix. The exGauss pdf is plotted Fig. 8.1 using values characteristic of wild data (from the CBB dataset). The exGauss pdf increases very quickly in practice with pointing data, due to the small standard deviation of the Gaussian deviation.

The parameters $\widehat{\Theta}$ of the exGauss model are estimated by MLE on the pointing datasets, as detailed in the Appendix. An example of such fits is given in Fig. 8.2. The front, given by (8.1) with β_0 and β_1 obtained through MLE corresponds well to its ideal representation described in 5.4.2 and represented Fig. 5.5.

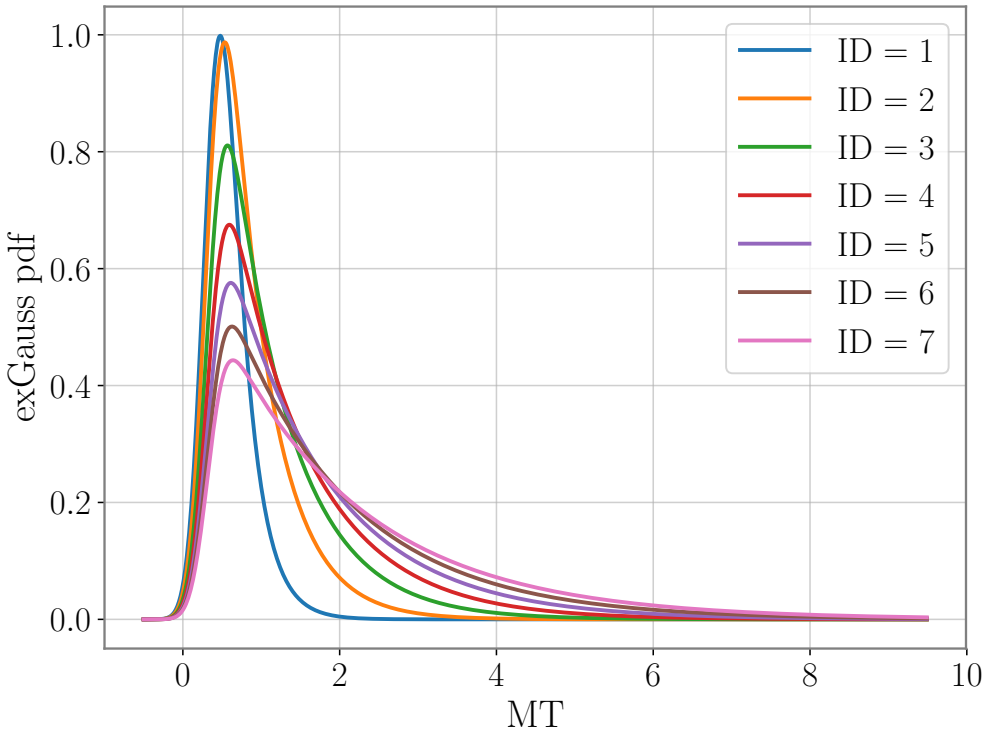

Figure 8.1: exGauss pdf plotted for $\mu = \beta_0 + \beta_1 \mathrm{ID}$, $\lambda = k/\mathrm{ID}$, σ^2, with $\beta_0 = 9\ 10^{-3}$, $\beta_1 = 7.9\ 10^{-2}$, $k = 3.876$, $\sigma^2 = 3\ 10^{-2}$, for different ID levels.

How the model actually fits the entire scatter plot is harder to see. A particular visualization can be useful here, see Fig. 8.3. We first bin the scatter plot along the ID axis to reduce our problem to a 1-D problem. Then, we can visualize the fit of the model by three equivalent manners: a plot of the empirical CDF compared with the theoretical CDF (Fig. 8.3, top right panel), a quantile-quantile (QQ) plot (Fig. 8.3, bottom left panel) or a histogram (Fig. 8.3, bottom right panel). As the QQ plot is ideally a straight line, we can represent the QQ plots of all bins on a single figure, see Fig. 8.4, by rotating the QQ plots by $\frac{\pi}{4}$, so that it should ideally be aligned vertically.

We plotted all full QQ plots when enough data was available (we plotted centile-centile plots, hence more than 100 points are required) for all datasets of Fig. 8.2, see Fig. 8.5. The fact that many QQ-plots are curved indicate a mismatch with the shape of the distribution that is fitted. As seen in Fig. 8.3, a QQ-plot that is on the left of the identity line indicates that the model underestimates the chance of an event, while a QQ-plot on the right of the identity line indicates that the model overestimates the chance of an event. Hence, a curvature to the from left to right or from right to left is likely due to a skewness mismatch between the model and the data, caused by a

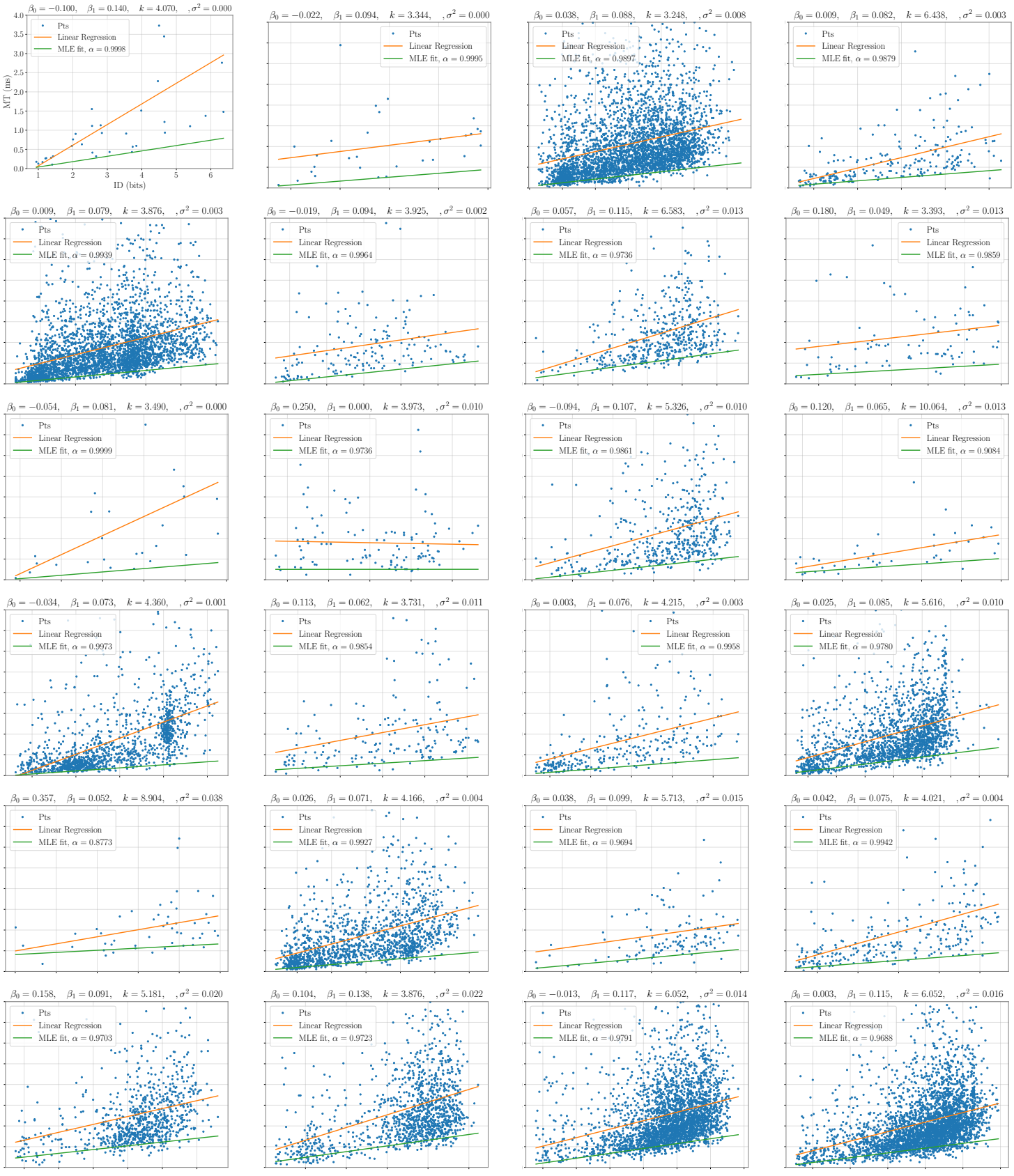

Figure 8.2: MLE fits with the exGauss model compared to the linear regression for one participant of the CBB dataset[14]. Each plot is the wild pointing data for one specific widget.

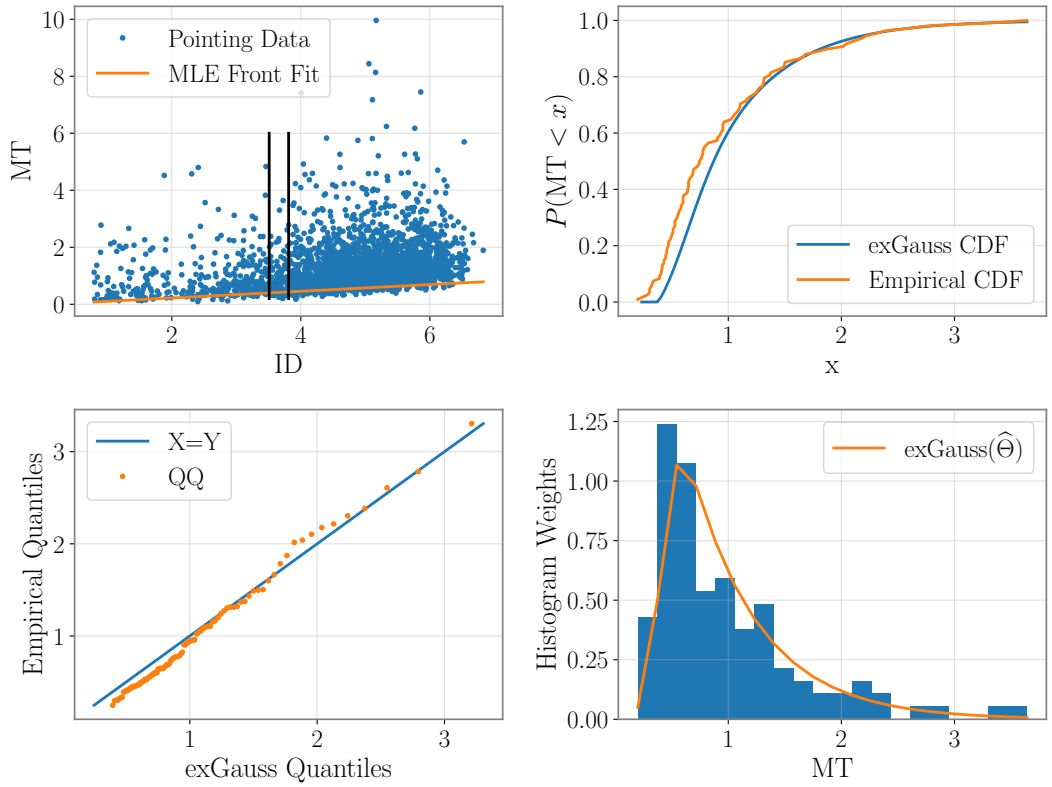

Figure 8.3: Top Left Panel: Movement time and MLE Front fit. The black bars indicate a bin, whose data is used in the other three panels. Top Right: Empirical CDF and exGauss CDF for the bin isolated in the top left panel. Bottom Left: Quantile-Quantile plot for the bin isolated in the top left panel. Bottom Right: Histogram for the bin isolated in the top left panel and exGauss pdf.

tail that does not decrease at the correct rate. In fact, the skewness of the exGauss model cannot be adjusted, or atleast not when the value of σ^2 is small. In that case, the skewness of the exGauss is approximately the skewness of the exponential law, which is constant and equal to 2. This notion of small σ can be defined precisely by noting that the skewness of the exGauss function is given by

$$\mathrm{Skew}(\mathrm{exG}) = \frac{2}{\sigma^3\lambda^3}\left(1 + (\sigma\lambda)^{-2}\right)^{-3/2}, \tag{8.16}$$

which reduces to 2 for $\sigma\lambda << 1$. In our fits, this condition indeed holds.

Hence a more complex model that includes an extra degree of freedom, relating to the shape of the distribution should provide a better fit. An idea is to consider the family from which we chose the exponential distribution initially, i.e. the gamma distribution. However, this comes at the cost of having an extra parameter to interpret and is left for future work.

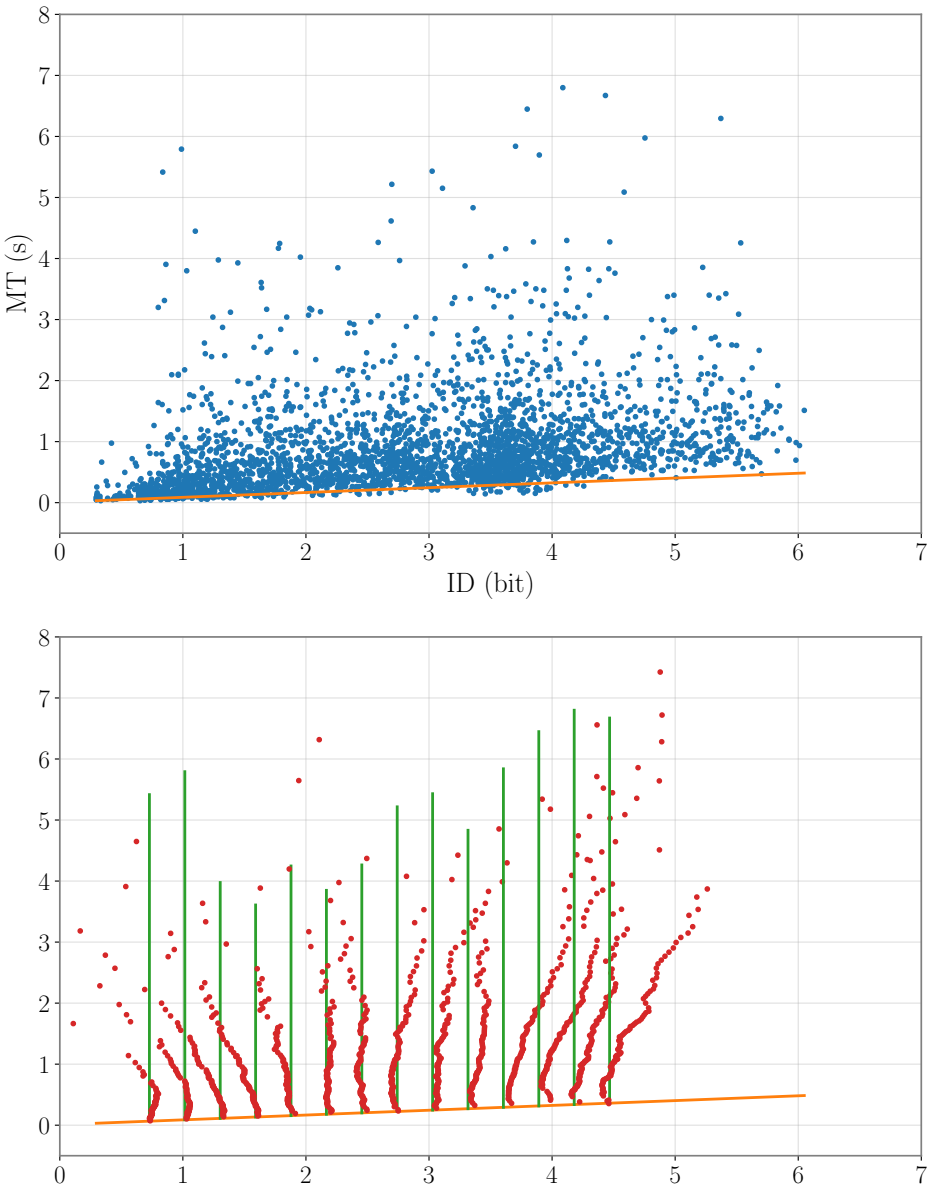

Figure 8.4: Top Panel: Movement time and MLE Front fit. Bottom panel: Front fit (orange), ideal QQ plots for each bin (green), corresponding empirical QQ plots (red). Axes of the bottom panel are equal to those of the top panel.

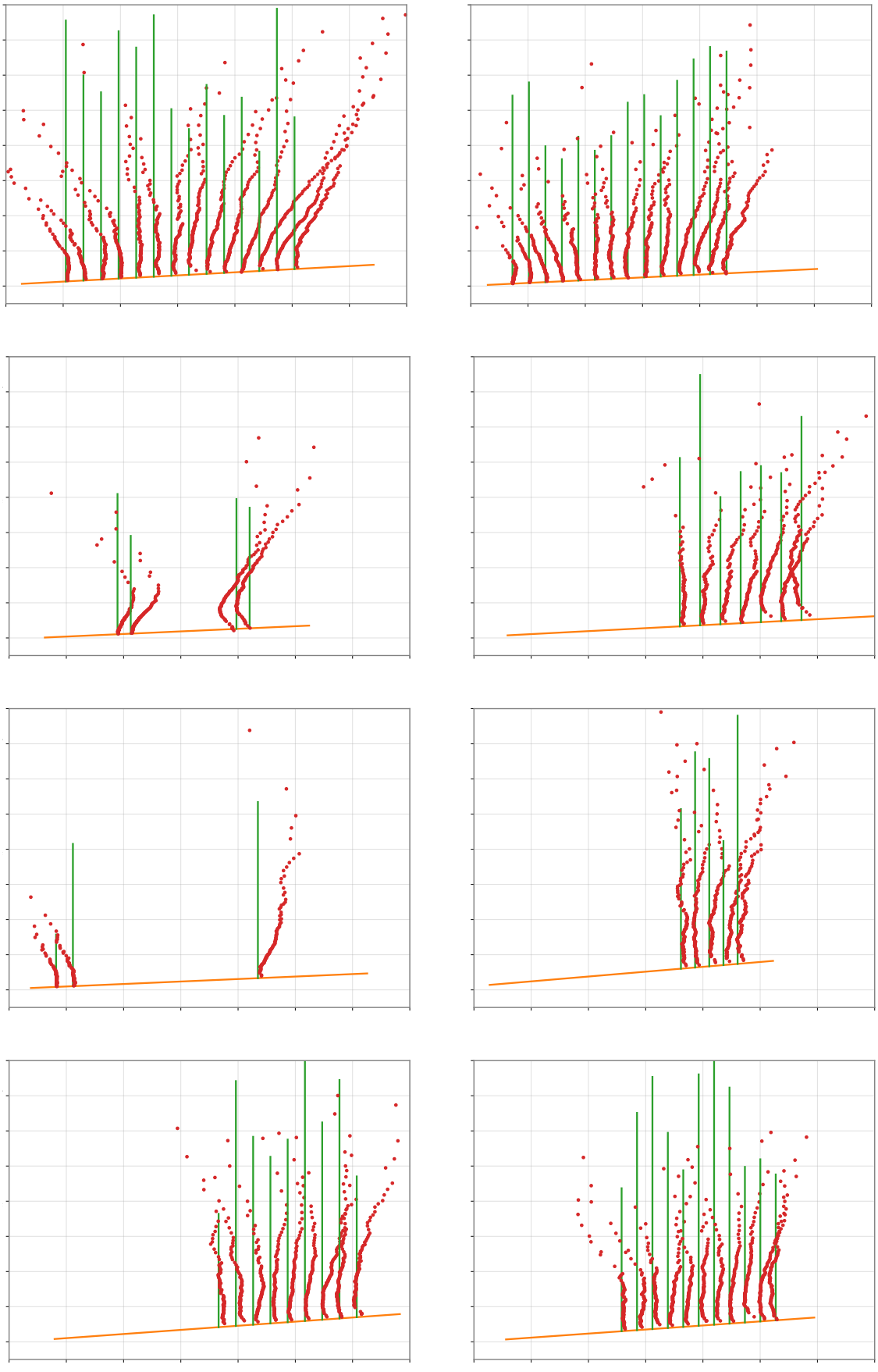

Figure 8.5: Empirical QQ plots for the 8 datasets used in Fig. 8.2 that contain the most data. Axes are identical to those of Fig. 8.4.

8.1.3 Consistency with Fitts' Law

The exGauss model (8.12) is consistent with Fitts' law. To see this, let us compute the average movement time $\mathbb{E}(y)$

$$\mathbb{E}(y) = \mathbb{E}(A) + \mathbb{E}(\varepsilon), \tag{8.17}$$

by linearity of the mathematical expectation. The average value of an exponential law of parameter k/x is simply x/k, while the average of the Gaussian is simply the mean parameter. Hence

$$\mathbb{E}(y) = x/k + \beta_0 + \beta_1\, x = \beta_0 + (\beta_1 + 1/k)\; x, \tag{8.18}$$

which is compatible with Fitts' law. In fact the result does not only hold for the average movement time, but also approximatively for the median, as well as any quantile. The quantile distribution is not directly accessible as the CDF of the exGauss function is complicated to invert. However, a simple approximation comes from considering the exponential random variable only, as the Gaussian deviation has a small variance. The quantile function $Q(F)$ for the exponential law of parameter x/k is

$$Q(F) = -\log(1-F) \cdot x/k. \tag{8.19}$$

Thus, for each quantile F, the following holds approximately

$$y(F) = \beta_0 + \big(\beta_1 - \log(1-F)/k\big)\; x. \tag{8.20}$$

The regression from quantiles was computed by Chapuis et al. [14], who indeed found good fits with a Fitts model for centiles between the 20^{th} and the 90^{th}, although the slopes they computed were concave in F, while our formula is convex in F ($\log$ being concave, $-\log$ is convex). Furthermore, Chapuis et al. also report changes in intercept during the different quantiles fit, which cannot be explained by 8.20 that predicts a constant intercept β_0 (see 8.2.2 for further analysis).

8.2 Application to Empirical data

The exGauss model can be applied to data acquired from the very controlled setup of the laboratory to an "in the wild" setting. In this section, the model is examined in these two conditions, as well as in the intermediate setup of web-based experiments. There, the task is as unnatural as in the laboratory, but the participant is not incited to perform as much as in a controlled experiment.

8.2.1 Controlled Data

The exGauss model was fitted on data from the GFA dataset. A couple of example fits are shown Fig. 8.6. In all cases, there seems to be very little difference in the

intercepts, while the slopes are slightly lower. the values for k are above 15, much larger than the fits found in the previous section for wild data. This is expected, as larger values of k indicate a high rate for the exponential and the participant is often close to the front of performance i.e. the participant is performing in the neighborhood of its best performance. Average value and standard deviation for α are respectively 0.922 and $6.6\ 10^{-2}$.

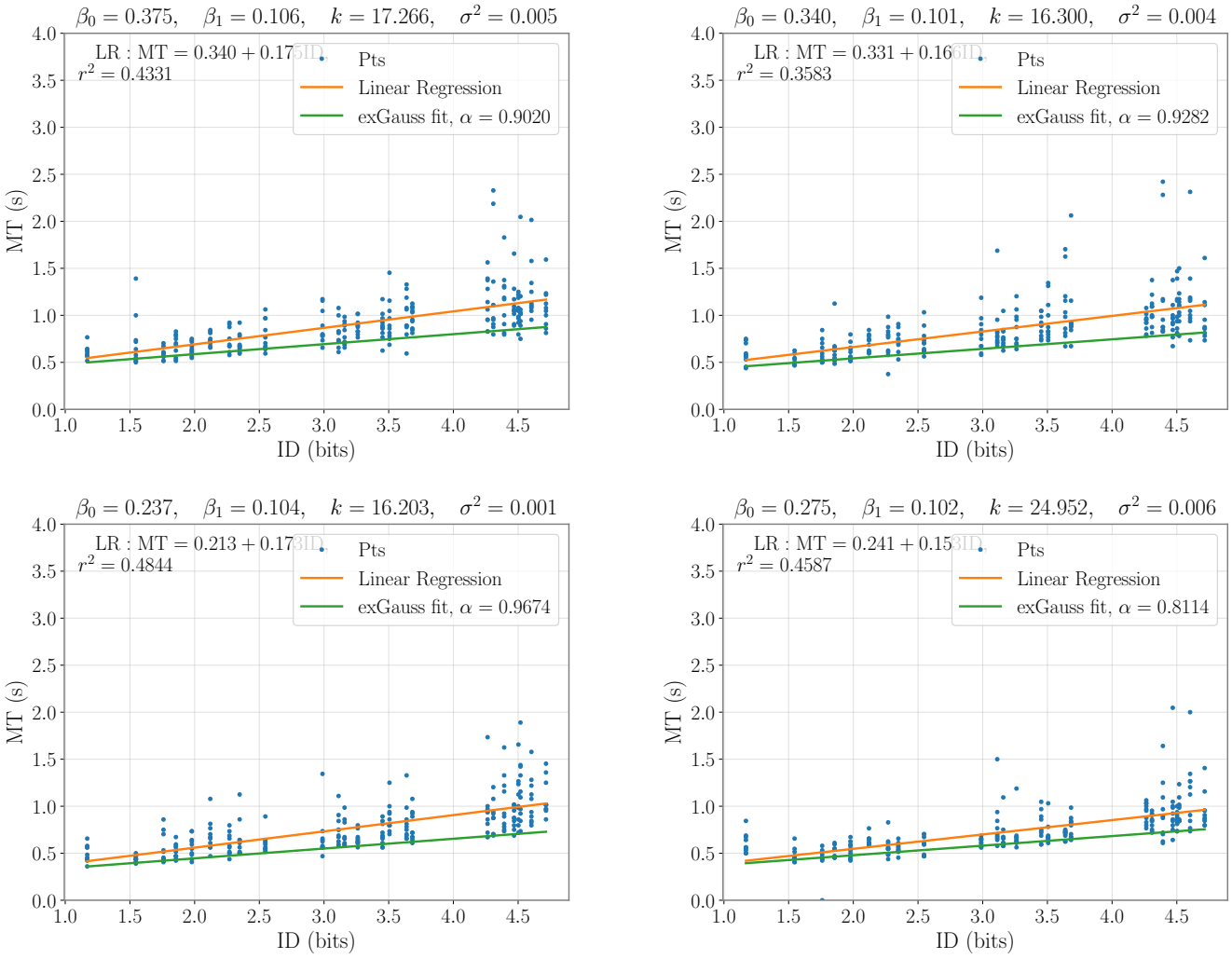

Figure 8.6: 4 randomly chosen linear regression (LR) and exGauss fits performed on the GFA dataset. The LR parameters are displayed in the plot and the LR fit is plotted in orange, while the exGauss parameters are displayed above the plot, and the MLE exGauss fit is plotted in green.

These observations hold for the entire dataset, see Fig. 8.7 and Table 8.1. The value of the intercept barely changes (on average it increases by 2 ms), in line with (8.18). On

the other hand, the slope decreases on average by 56 ms/bit, which is approximately accounted for by k as predicted by (8.18) ($1/k = 54$ ms/bit). The prediction (8.18) is further verified on the bottom right panel of Fig. 8.7, as the scatter plot aligns nicely along the identity line ($X = Y$).

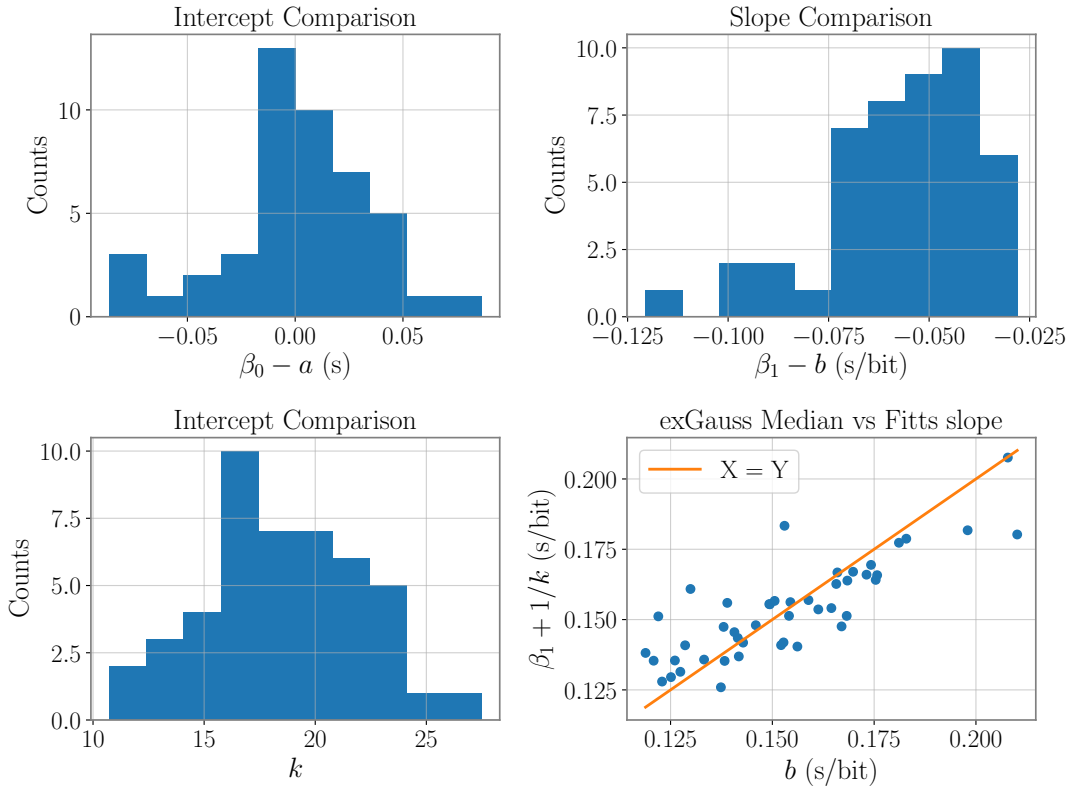

Figure 8.7: Summary of the GFA dataset. Top Left panel: Histogram for the differences in intercept between the exGauss and linear regression model. Top Right: Histogram for the differences in slope between the exGauss and linear regression model. Bottom Left panel: Histogram for the k parameter in the exGauss model. Bottom Right panel: Comparison between the mean of the exGauss model and the slope of the linear regression model.

Table 8.1: Parameter summary for exGauss fit on the GFA dataset.

	$\beta_0 - a$ (s)	$\beta_1 - b$ (s/bit)	k	σ^2 (s)
mean	$2 \cdot 10^{-3}$	$-5.6 \cdot 10^{-2}$	18.5	$2.9 \cdot 10^{-3}$
std. dev.	$3.6 \cdot 10^{-2}$	$1.9 \cdot 10^{-2}$	3.7	$2.7 \cdot 10^{-3}$

8.2.2 Data Acquired "in the Wild"

The exGauss fits for the wild CBB dataset were already given Fig. 8.2. Comparisons between the exGauss model and the linear regression fit are given Fig. 8.8 and Table 8.2. The intercept and slopes are reduced respectively by about 100 ms and 100 ms/bit from the linear regression fit to the exGauss fit. The value of k is much lower than with the controlled data (average $k = 8$, about half of the value for controlled pointing. Average value and standard deviation for α are respectively 0.901 and $1.26\ 10^{-1}$.

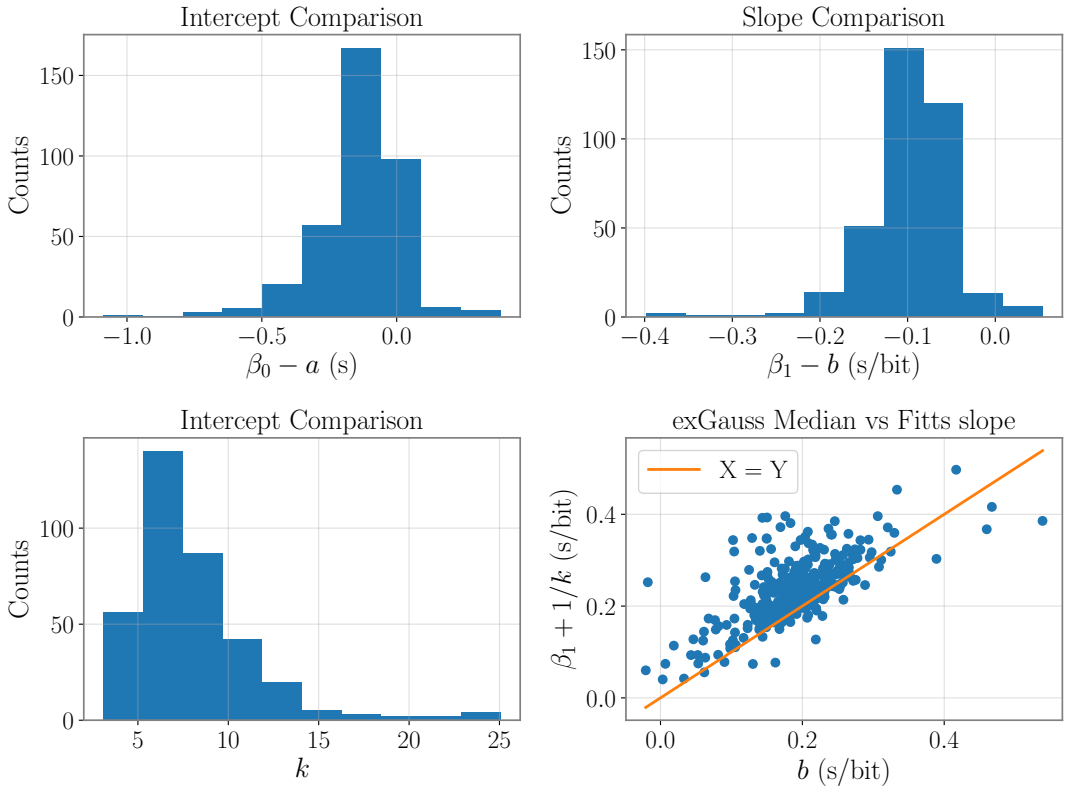

Figure 8.8: Summary of the CBB dataset. Top Left panel: Histogram for the differences in intercept between the exGauss and linear regression model. Top Right: Histogram for the differences in slope between the exGauss and linear regression model. Bottom Left panel: Histogram for the k parameter in the exGauss model. Bottom Right panel: Comparison between the mean of the exGauss model and the slope of the linear regression model.

Table 8.2: Parameter summary for exGauss fit on the CBB dataset.

	$\beta_0 - a$ (s)	$\beta_1 - b$ (s/bit)	k	σ^2 (s)
mean	$-1.3 \cdot 10^{-1}$	$-9.7 \cdot 10^{-2}$	8	$2.1 \cdot 10^{-2}$
std. dev.	$1.5 \cdot 10^{-1}$	$4.9 \cdot 10^{-2}$	3.4	$2.1 \cdot 10^{-2}$

8.2.3 Data Acquired in a Web-based Experiment

In this subsection we use the data from GFA dataset that was gathered in a web-based experiment; the device used by the participants are either the mouse, the trackball, the trackpad, the trackpoint, or unknown. The exGauss and linear regression fits are given for the datasets pertaining to the 4 identified devices see Fig. 8.9. The dataset does not provide information other than the device used by the participant, and so the datasets can not be separated into subsets other than the 4 already represented, we will thus not produce the same analysis as with the two previous datasets.

Levels of k are in between those obtained for controlled and wild experiments, as expected (average value $k = 9.834$). The values for α are below those found for controlled and wild data. This is likely due to the fact that each web-based dataset is the aggregation of many different users (e.g. about 1500 different users for the mouse), while the controlled dataset separates data by user, and the wild dataset even separates the data by targets. Average value and standard deviation for α are respectively 0.844 and $6.28\ 10^{-2}$.

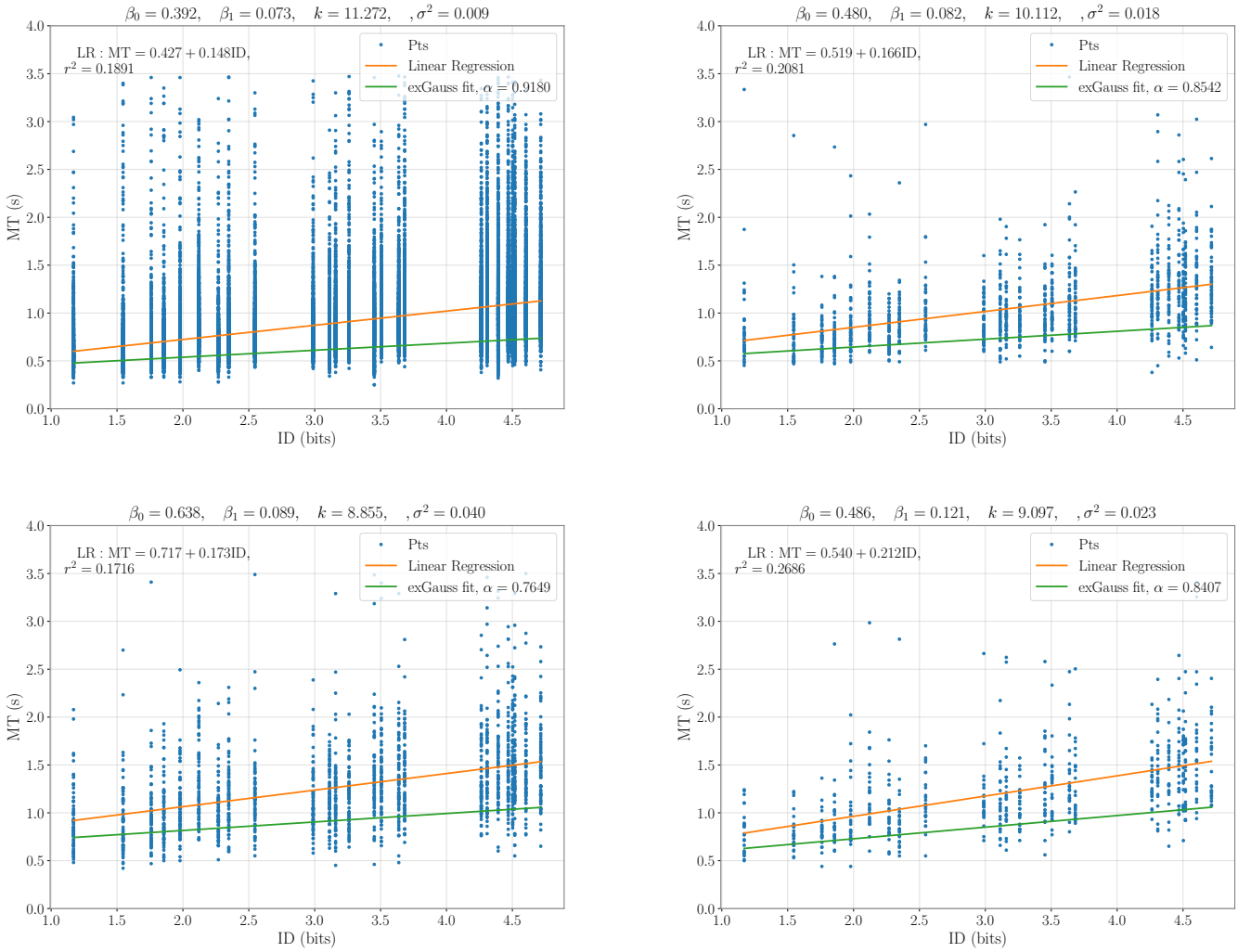

Figure 8.9: exGauss and linear regression fits for GFA web-based experiment with 4 different devices (Top left: mouse, Top right: trackball, Bottom left: trackpad, Bottom right: trackpoint).

8.2.4 Discussion

The exGauss model suggests using β_0 and β_1 rather than a and b, and provides two new parameters: k and σ^2.

β_0 versus a While the front of performance does not result in a change of intercept for the controlled data, as predicted by (8.18), there is a 100 ms decrease on average from a to β_0 for the wild dataset. The β_0 obtained for the wild dataset is about 100 ms (0.102 s on average, standard deviation 0.127 s), which is close to typical values

observed in controlled experiments in the literature [142]. A tentative reason to explain this decrease that occurs only in the wild condition is that in a controlled experiment there are typically very few points acquired compared to a wild study, hence the front might only sparsely be described. More investigation is needed at this point to formulate any serious hypothesis. The web-based experiment also sees a decrease of about 50 ms, from a to β_0 (average: -0.052, standard deviation: 0.020).

β_1 versus b In all cases, the slopes decrease from b to β_0, in line with (8.18), from 50 ms/bit for the controlled case to 100 ms/bit in the wild case. Interestingly, all values for β_1 are about 100 ms/bit. We have respectively, for the wild, controlled and web-based experiment values of 0.092 (std.dev. 0.039), 0.097 (std.dev. 0.013) and 0.091 (std.dev. 0.021). Hence, it seems that β_1, contrary to b, stays relatively constant (at about 100 ms/bit) across experimental conditions.

k Values of k increase when the experiment is more and more controlled as expected. Values of k for the web-based experiment are very close to those of the wild condition, suggesting that web-based experiments, even though they usually implement an unrealistic task, may prove useful to understand real life interactions.

α Until now, the unspoken rule in our analysis has always been that on one side were results from the wild condition, on the other from the controlled condition, and somewhere in the middle were the results from the web-based experiment. This seemed reasonable, as the web-based experiment seems to make the compromise between the stereotyped tasks of the laboratory and the natural environment of the wild study. With α, matters change, as it is actually lowest for the web-based experiment. Visually, the web-based dataset is also the one that offers the least satisfying fits. My hypothesis is that the web-based dataset is the most aggregated one. This underlines an empirical difficulty with the front of performance approach, which can be disrupted if there are too many fast outliers. More work can probably lead to more robust techniques to determine the front.

8.3 Appendix

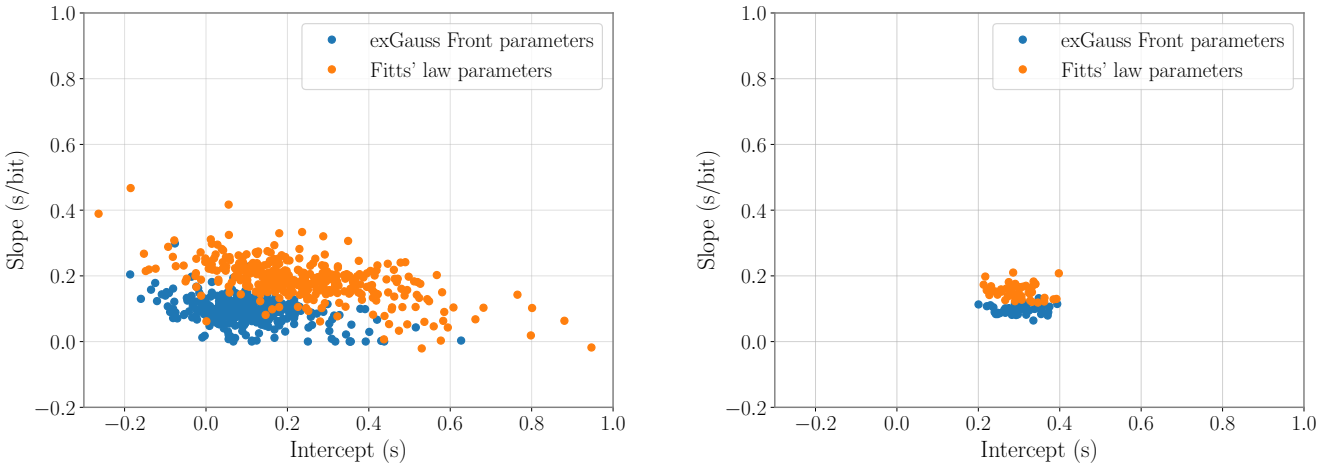

Figure 8.10: Comparison of the exGauss front parameters versus the linear regression Fitts' law parameters for the wild (left), and controlled (right) datasets.

8.3.1 Principle of Maximum Likelihood Estimation (MLE)

The framework for MLE is as follows. We are given a model (a pdf) f which takes argument Θ (usually a vector) and an admissible range for Θ. We are also given the observed data (dependent and independent variable). The goal of MLE is to find the parameters of the model which makes the data most probable. Informally, the question that MLE tries to answer is: "what are the parameters that, given the model f, make the observed data the most probable". It does so by computing the probability of the observed data, called the likelihood $\mathcal{L}$ of a model and then maximizing the likelihood over Θ.

8.3.2 exGauss Distribution MLE

A random variable E that follows an exGauss distribution has the following pdf

$$p_E(y|\mu,\sigma^2,\lambda) = \lambda/2\exp\big(\lambda/2(2\mu+\lambda\sigma^2-2y)\big)\operatorname{erfc}\big(\frac{\mu+\lambda\sigma^2-y}{\sqrt{2\sigma^2}}\big), \tag{8.21}$$

where erfc is the complimentary Gaussian error function.

In our model, movement time y is obtained as the sum of A and ε (8.7). Hence, its pdf is given by

$$p_Y(y) = \int_{\mathbb{R}} p_\varepsilon(u)\cdot p_A(y-u)du \tag{8.22}$$

$$= \int_{-\infty}^{y} p_\varepsilon(u)\cdot p_A(y-u)du, \tag{8.23}$$

where 8.22 is simply the convolution between the two pdf's and where the positive-definite property of A is exploited in 8.23. With the pdf's for A 8.13 and ε 8.14, we get

$$p_Y(y|\Theta, x) = k/x \exp(-ky/x) \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(u-\mu_1)^2}{2\sigma^2}\big) \exp(ku/x)du, \tag{8.24}$$

where $\mu_1 = \beta_0 + \beta_1\, x$. Let $\mu_2 = \mu_1 + k\sigma^2/x$; then (8.24) can be written more compactly as

$$p_Y(y) = k/x \exp(-ky/x) \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\big(-\frac{(u-\mu_2)^2}{2\sigma^2}\big) \exp\big(\frac{\mu_2^2 - \mu_1^2}{2\sigma^2}\big)du. \tag{8.25}$$

The last exponential in (8.25) does not depend on the integrand anymore and can be removed from the integral, and the remainder inside the integral is simply a Gaussian distribution, whose integral can be expressed using the Gaussian error function (see (3.29))

$$p_Y(y) = k/x \exp(-ky/x) \frac{1}{2} \left[1 + \text{erf}\ \big(\frac{y-\mu_2}{\sqrt{2\sigma^2}}\big)\right] \cdot \exp\big(\frac{\mu_2^2 - \mu_1^2}{2\sigma^2}\big), \tag{8.26}$$

which is an exGaussian distribution with $\mu = \beta_0 + \beta_1\, x$, $\sigma = \sigma$ and $\lambda = k/x$.

With N the total number of samples for y, the likelihood $\mathcal{L}(\Theta; y, x)$ is then obtained from $p_Y(y_1, y_2, \ldots, y_N) = \prod_{i=1}^{N} p_Y(y_i)$, as the samples are assumed independent. Usually it is simpler to consider the log-likelihood l and to remove constant terms which don't affect maximization.

$$l(\Theta; y, x) = \log \mathcal{L}(\Theta; y, x) \tag{8.27}$$

$$\propto \sum_{i=1}^{N} \left[\log k/x_i - ky_i/x_i + \log(1 + \text{erf}\ \big(\frac{y_i - \mu_2}{\sqrt{2\sigma^2}}\big)) + \frac{\mu_2^2 - \mu_1^2}{2\sigma^2}\right] \tag{8.28}$$

The optimal parameter vector $\widehat{\Theta}$ is found when maximizing the likelihood, or equivalently, the log-likelihood:

$$\widehat{\Theta} = \arg\max l(\Theta; y, x). \tag{8.29}$$

With such a model, it can be complicated to find a closed-form expression for $\widehat{\Theta}$, but numerical techniques can be applied effectively.

Chapter 9

FITTS 2: Empirical Validation

This chapter provides an empirical analysis of PVPs and effects of task parameters upon them. Since many pointing studies on Fitts' tapping task have been made, we have chosen to re-analyze two existing datasets:

9.1 Datasets

To validate FITTS 2 we used two 1-D datasets, namely the G-dataset and the PD-dataset (see Chapter 7). In both datasets, position is a one-dimensional real number. The raw data from each dataset was re-sampled closest to the average sampling frequency, and individual movements were extracted and synchronized as described in 6.1.1. For the G-dataset, 16 participants produced movements for each of the five conditions five times, providing 80 different PVPs. For the PD-dataset, 12 participants produced movements for 8 (D,W) pairs, providing 96 PVPs. Outliers were not removed because this would imply relying on arbitrary heuristics for which we found no satisfying method. We were nonetheless able to find compelling evidence to support our claims.

9.2 PVP unimodality

We asserted unimodality on both datasets by verifying that only one sign change occurred in the derivative of the standard deviation $\sigma(t)$ after the peak of variance[1]. We considered that the sign of the derivative changes when it crosses a threshold level of $\pm 2.5\%$ of the maximum value of the derivative. Fig. 9.1 displays an empirical PVP, consistent with the theoretical profile of Fig. 6.2, and shows its derivative in orange. There is just one sign change, when $\sigma(t)$ is at a maximum, which indicates unimodality.

We conducted this analysis on all PVPs from both datasets: 94% (75 out of 80) of PVPs were found unimodal in the G-dataset and 92% (88 out of 96) in the PD-dataset.

[1] It is equivalent to investigate standard deviation $\sigma(t)$ or variance $\sigma^2(t)$, but as seen below, the values of standard deviation are easier to interpret.

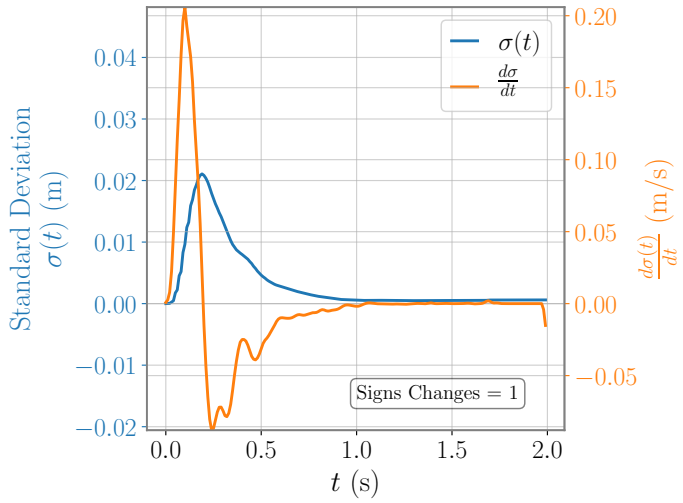

Figure 9.1: In blue: Standard deviation profile $\sigma(t)$ for Participant X of the G-dataset, performing under condition 3 (balanced speed/accuracy). In orange: Derivative of the standard deviation profile with respect to time.

The Hartigan & Hartigan dip test was conducted for unimodality [73] on the G-dataset, where the null hypothesis is that the distribution is unimodal. We found that 6 profiles ($p < .05$) were significantly non-unimodal out of the 80 profiles. The average p-value is just above 0.8, implying strong evidence for unimodality. Evidence for the PD-dataset was less conclusive using the Hartigan & Hartigan dip test. As the task is reciprocal, the initial variance is not null, hence the second phase displays quite often a small decrease in variance before assuming the unimodal profile, see Fig. 9.2. Hence the overall support for unimodality is strong.

9.3 Effects of D, ID, W and instruction on τ and D_τ

As explained in 2, the onset of the movement is primarily affected by D, while the later stages are primarily affected by W. Also, Fitts' law predicts that most of the endpoints time's variability can be explained by ID levels. Therefore, we investigated the effect of D, ID and W on τ, the time of maximum variance, as well as D_τ, the distance covered at maximum variance. The effects of instruction on τ and D_τ, based on the G-dataset, are also reported for completeness,

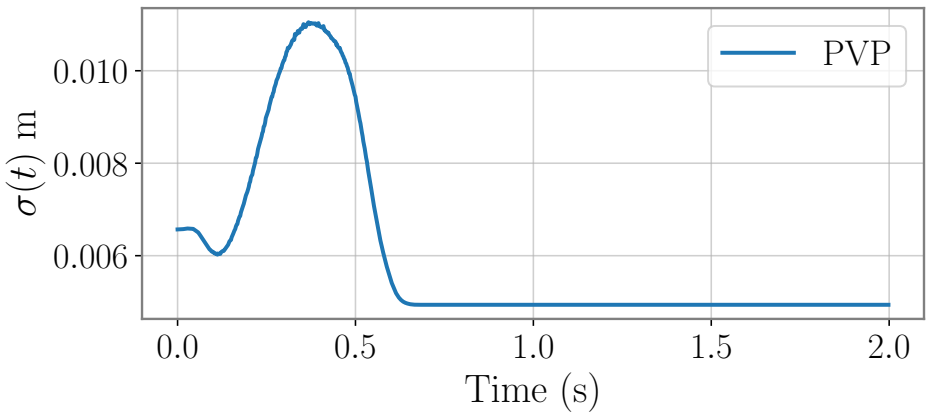

Figure 9.2: PVP for the reciprocal task (PD-dataset) for $\mathrm{ID} = 2$.

9.3.1 Effects of D, ID and W on the time instant τ of maximum variance

These effects are summarized in Fig. 9.3. Grand mean for τ is $\overline{\tau} = 0.312$ s, while average τ per condition ranges from 0.283 to 0.357 ms. On average, τ increases slightly with ID. Effects on τ are very small compared with the increase in total movement time (represented in black diamonds in Fig. 9.3), ranging from 0.49 s to more than 1.5 s.

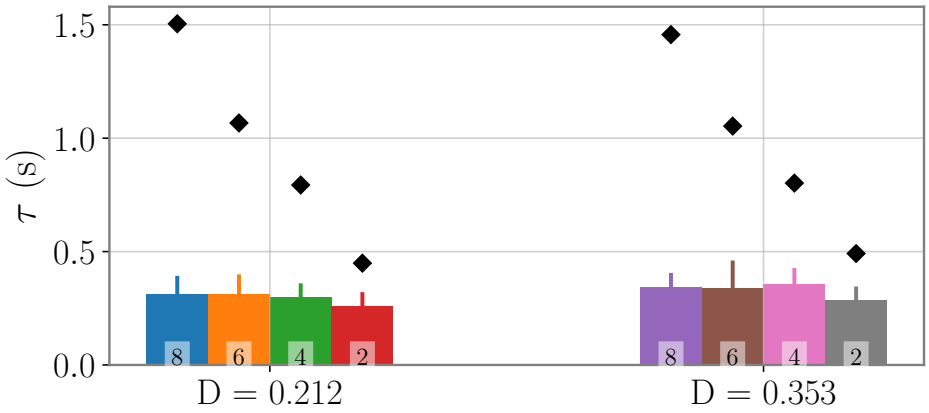

Figure 9.3: Effects of D and W on τ for the PD-dataset. The bars are grouped by D condition, with the 4 bars on the left corresponding to $\mathrm{D} = 0.212$ m and the 4 bars on the right corresponding to $\mathrm{D} = 0.353$ m. Each bar is labeled with its corresponding level of ID. For comparison, total movement time is represented with black diamonds with the same scale.

Table 9.1: Two-way ANOVA for effects of D and ID on τ.

	F	p	η^2
D	$F(1,88) = 4.779$	0.031	0.047
ID	$F(3,88) = 2.831$	0.043	0.083
DxID	$F(3,88) = 0.236$	0.871	0.007

This is confirmed by a two-way ANOVA, see Table 9.1. The effects of D and ID are statistically significant at $\alpha = 0.05$; however the effect size is small, indicating that less than 10% (resp. 5%) of the variance is explained by factor ID (resp. D). In comparison, a two-way ANOVA for effect of D and ID on total movement time results in an effect size of more than 70% of ID on movement time. Therefore, ANOVAs confirm the visual impression that the effects of the factors on τ are small compared to the effects on movement time.

The coefficient of determination between τ and W was found as $\tau = 0.327 - 0.0005 \times \mathrm{W}$, $r^2 = 0.0557$, $p = 0.021$. Although the slope is statistically significant at $\alpha = 0.05$, it is almost horizontal and r^2 is very close to null, showing that W has very little effect on τ as well.

9.3.2 Effects of D, ID and W on the distance traveled D_τ at time τ

The effects are summarized in Fig. 9.4. ID levels have little or no effect on D_τ Higher levels of D lead to larger D_τ:

- For D = 0.212, D_τ ranges from 0.127 (for $\mathrm{ID} = 4$) to 0.147 m (for $\mathrm{ID} = 6$). The average D_τ over all ID levels is 0.138, representing 65% of D.
- For D = 0.353, D_τ ranges from 0.217 (for $\mathrm{ID} = 2$) to 0.238 m (for $\mathrm{ID} = 8$). The average D_τ over all ID levels is 0.228, which also represents 65% of D.

A two-way ANOVA reported in Table 9.2 confirms our analysis. Only D is statistically significant at $\alpha = 0.05$, with very strong effect size ($\eta^2 > 60\%$), while both ID and the interaction $\mathrm{ID} \times \mathrm{D}$ have no significant effect on D_τ. Linear regression between W and D_τ gives $\mathrm{D}_\tau = 0.0001 \times \mathrm{W} + 0.1807$, $r^2 = 0.0024$, $p = 0.633$, which shows no effect of W on D_τ.

	F	p	η^2
D	$F(1,88) = 143.686$	< 0.001	0.609
ID	$F(3,88) = 1.241$	0.300	0.016
DxID	$F(3,88) = 0.212$	0.888	0.003

Table 9.2: Two-way ANOVA for effects of D and ID on D_τ.

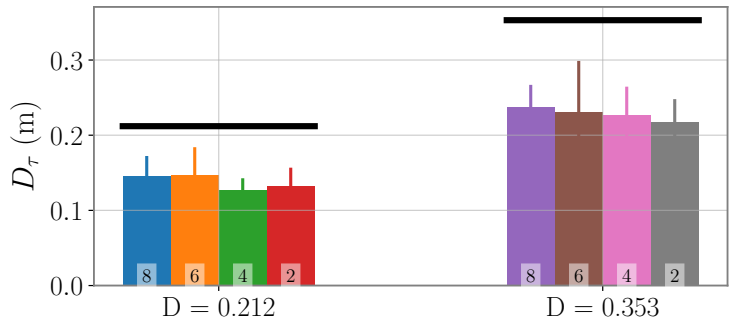

Figure 9.4: Effects of D and W on τ for the PD-dataset. The bars are grouped by D condition, with the 4 bars on the left corresponding to $D = 0.212$ m and the 4 bars on the right corresponding to $D = 0.353$ m. Each bar is labeled with the corresponding value of ID. For comparison, the level of factor D is represented in thick black horizontal lines.

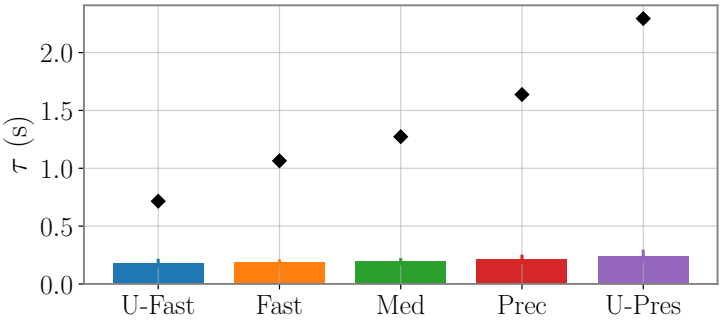

Figure 9.5: τ against instructions. Black diamonds represent total movement time.

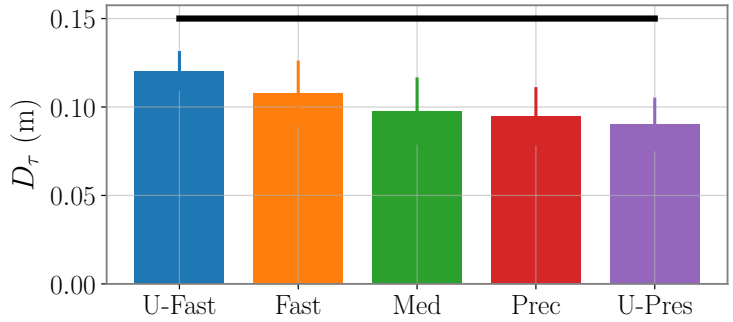

Figure 9.6: D_τ against instructions. The distance to be reached $D = 0.15$ m is represented with a thick black line.

9.3.3 Effect of instructions on τ and D_τ

These are summarized in Figs. 9.5 and 9.6.

- Effect on τ: Grand mean for τ is $\overline{\tau} = 0.201$ s; τ ranges from $\tau = 0.180$ s for the ultra fast condition to $\tau = 0.240$ s for the ultra precise condition. Instructions to emphasize precision lead to higher values of τ. A one-way ANOVA on the factor instruction gives $F(4, 75) = 6.200$, $p < 0.001$, $\eta^2 = 0.249$, which shows a statistically significant, moderate effect of instruction on τ. In comparison, the effect of instruction on total movement time, displayed in black diamonds in Fig. 9.5, is much stronger ($\eta^2 = 0.491$).
- Effect on D_τ: grand mean for D_τ is 0.102 m; D_τ ranges from $\mathrm{D}_\tau = 0.120$ m for the ultra fast condition to $\tau = 0.090$ m for the ultra precise condition. Instructions to emphasize speed lead to larger values of D_τ. A one-way ANOVA on D_τ with factor instruction gives $F(4, 75) = 8.579$, $p < 0.001$, $\eta^2 = 0.314$, which shows a statistically significant, moderate effect on D_τ.

9.3.4 Summary

To summarize:

1. The duration of the first (variance-increasing) phase of the PVP is invariant to changes in D, W, ID and, to a lesser extent, to instructions. Although the effects on these factors are sometimes statistically significant, they are always very small compared to those on total movement time (time differences ≤ 100 ms vs. ≥ 1 s on total movement time in the PD-dataset; ≤ 60 ms vs. ≥ 1.5 s in the G-dataset). In the PD-dataset, differences of τ induced by varying the factors do not exceed 100 ms, whereas differences in movement time can exceed 1 s. In the G-dataset, the contrast is even greater, with difference in τ values of at most 60 ms, compared to differences in movement time that can reach more than 1.5 s.
2. The distance covered during that first phase is about two-thirds of the total distance to be covered ($\mathrm{D}_\tau \approx 0.65\ \mathrm{D}$ for both levels of D in the PD-dataset; $\mathrm{D}_\tau \approx 0.66\ \mathrm{D}$ on average on conditions (2)–(4) in the G-dataset). In the PD-dataset we found that both levels of D induce $\mathrm{D}_\tau = 0.65\ \mathrm{D}$. In the G-dataset, averaging the D_τ for conditions (2)–(4) (those which correspond to instructions given in the P-dataset) results in $\mathrm{D}_\tau = 0.66\ \mathrm{D}$.

Trajectories in the tapping task can thus be described by (1) a first variance-increasing phase of *constant duration* that lasts about 200 to 300 ms, where about $2/3$ of the distance is covered, followed by (2) a second variance-decreasing phase to meet the accuracy constraints. This simple two-phase description is a good match with two-component models of Woodworth [167] and Elliott et al. [34, 33].

Considering τ as constant greatly simplifies considerations since the effect of D, ID, W and instructions on τ lead to differences whose magnitude matches the within

groups variability. For most movements that require precision, not only does the second phase *dominate* movement time, as expected in 6, but it also induces Fitts' law as the first phase has constant duration.

In the G-dataset, τ is about 200 ms; in the PD-dataset, τ is about 300 ms. This value should be compared to the constant movement time ($\mathrm{MT} \simeq 190$ ms) reported by Crossman & Goodeve [22] for small values of ID (the dot-dashed green line in Fig. 3.2). A plausible explanation is that very short movements are essentially achieved without second phase. Hence, the more precise the movement, the more the second phase dominates movement time and the more our model is appropriate.

9.4 Link with kinematics

The previous analysis revealed that the effect of the various task parameters have little to no effect on the average value of τ. However τ does correlate high with various kinematic markers (r^2 of about 0.5 for e.g., date of maximum speed, date of the first zero-crossing of speed). These two observations are not contradictory, and simply express that the kinematic markers appear roughly always in the same order. When movement time is longer, all the kinematic markers are simply shifted to the right.

An interesting observation is that τ is very close to the date of average minimal acceleration γ (maximal deceleration), see Fig. 9.7. The green line labeled $X = Y$ is the identity operator, i.e., points close to the orange line have maximum variance occurring simultaneously with the minimal acceleration. The points are in fact slightly underneath the line, meaning τ is always a bit smaller than γ, hence the maximum variance occurs just before reaching the minimum acceleration.

In typical acceleration profiles, there is a fast increasing portion followed by a fast decreasing portion; this initial surge is then followed by many fluctuations [141]. It is interesting that the maximum variance coincides with the end of the initial surge. This suggests that the first phase may well be modeled by a bang-bang model [109], where maximum torque is applied, followed by minimal torque [48]. From there, torque decreases and the final decrease of acceleration as well as its fluctuations thereafter correspond to the second phase, where reliable aiming takes place. This interpretation is consistent with the observation that visual regulations and compensatory adjustments occur primarily after rapid force impulses [16, 53]. The link between kinematics and variance is also supported by Khan et al. [84], who describes that the maximum variance occurs for the peak acceleration and marks a new phase in the so-called variability ratio, and by Zelaznik et al. [169], who notice that the time to peak positive acceleration essentially remains constant as MT changes.

9.5 Empirical Results for the Second Phase

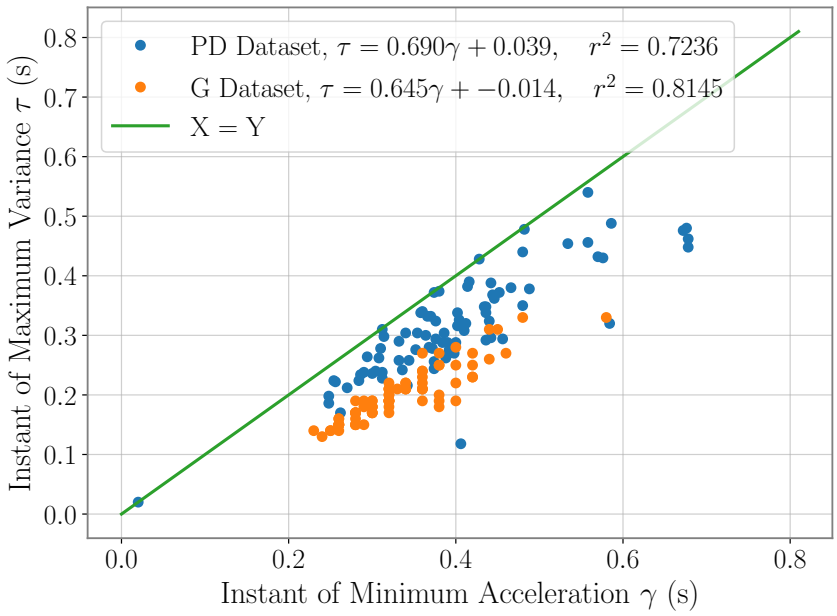

Figure 9.7: Instant of maximal variance (τ) plotted against the instant of minimal acceleration (γ). Each (γ,τ) point is extracted from one PVP. The green line, labeled $X = Y$ is the identity operator, i.e., points close to the orange line have maximum variance occurring simultaneously with the minimal acceleration.

9.5.1 Exponential Decrease of Standard Deviation

As shown in (6.21), the standard deviation is expected to decrease exponentially for optimal movements during the second phase. Assessing an exponential decrease is easily achieved within a log-lin scale. Within that representation, standard deviation plotted against time is expected to grow quickly during the first phase, and then decrease linearly with a slope $-C'$ during the second phase. When all movements have finished, the PVP should level off to a stationary phase, which lasts until the end of the observation window. Fig. 9.8 gives an example of a fit performed on the PVP. Standard deviation (in black) increases during the first 200 ms, then decreases quasi linearly up to about 950 ms (in blue) at rate $C' = 6.937$ bit/second. In the stationary phase (in orange) the standard deviation is about 0.5 mm.

We computed goodness of fit r^2's for all PVPs of the PD-dataset (resp. G-dataset) and obtained an average $r^2 = 0.963$ (resp. 0.971). The average C' found was $C' = 6.25$ (resp. 5.31) bit/s. Remarkably, such values are very close to the typical 5 bit/s throughput for a Fitts' law experiment as reported in [142]. This indicates a good consistency between our novel approach based on PVPs and the traditional approach based on total movement times. To further verify the consistency with traditional Fitts' law, we define movement time as the time when the linear fit of the second phase intercepts the fit of the stationary phase, see Fig.9.9. From there the typical plot of Fitts' law is eas-

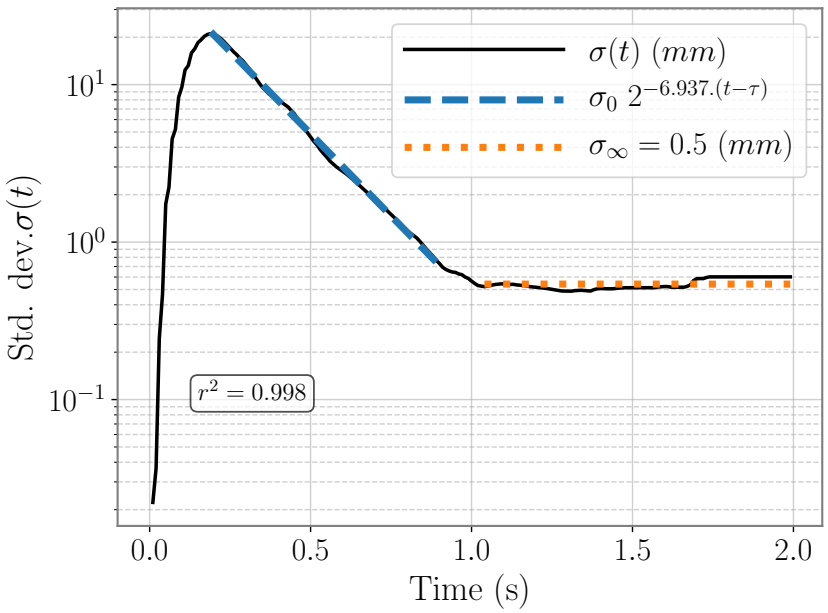

Figure 9.8: PVP for a participant of the G-dataset, performing under conditions of balanced speed and accuracy. The blue dashed line is the fit for the second phase ($r^2 = 0.998$); the orange dotted line is the fit for the stationary phase.

ily obtained, see Fig. 9.10, and results obtained this way are comparable to traditional Fitts' law analysis.

We conclude this part by illustrating the PVPs for one participant of the PD-dataset performing under all conditions (Fig. 9.11).

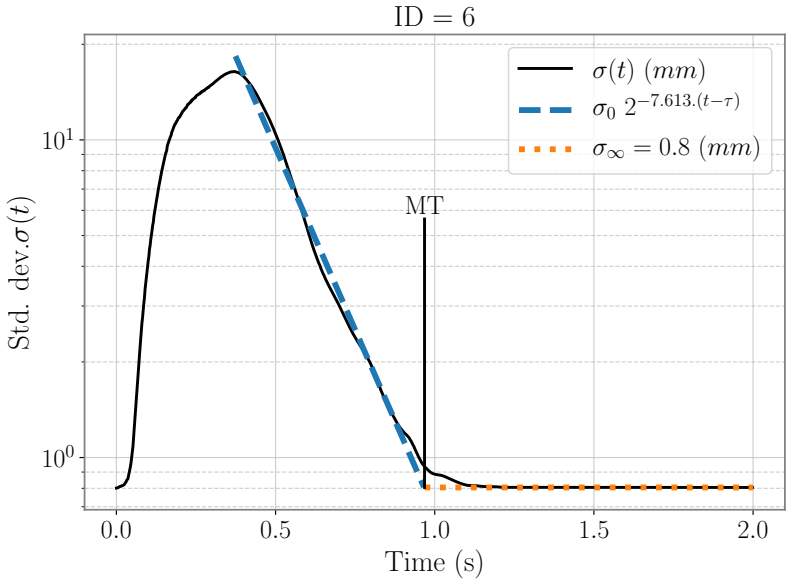

Figure 9.9: Defining MT from the PVP.

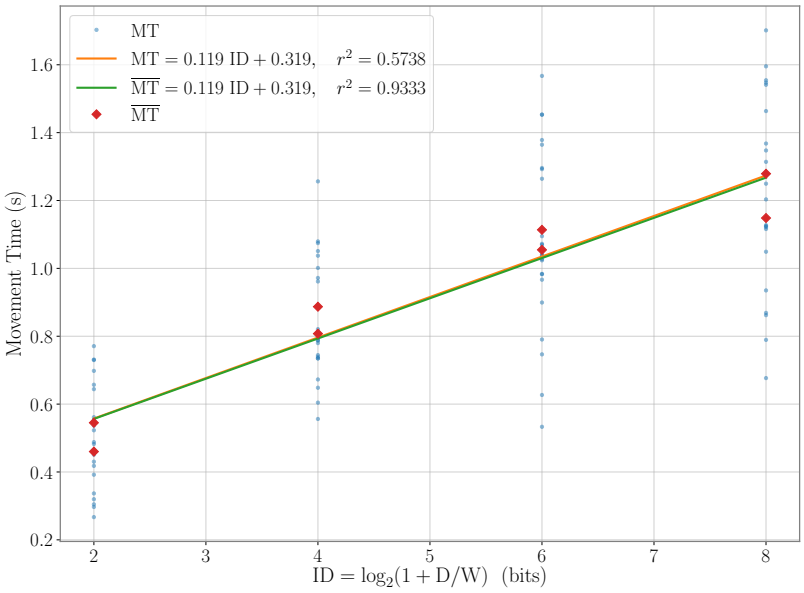

Figure 9.10: Movement time pre-average (MT) and post-average ($\overline{\mathrm{MT}}$) plotted against ID. The fitted lines are obtained by linear regression.

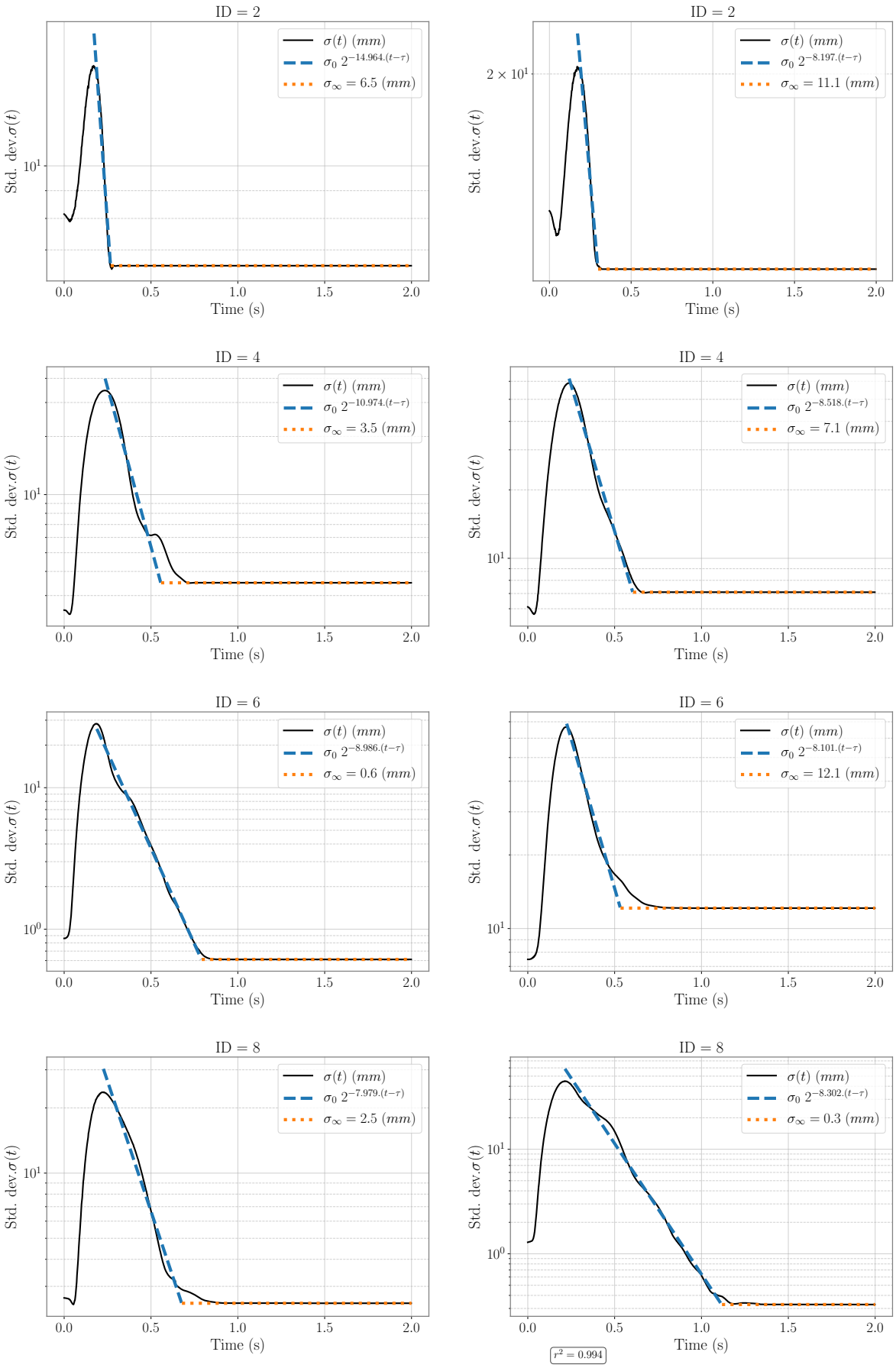

Figure 9.11: PVPs from participant 11, for all 8 conditions. On the left, the condition is $\mathrm{D} = 0.212$ m, on the right the condition is $\mathrm{D} = 0.353$ m. From top to bottom, the ID level goes from 2 to 8.

9.5.2 Effects of Task Parameters and Instructions

As the value of C' is determined by the transmission channel *only*, it is expected that varying the parameters D and ID leave C' mostly unchanged. In the example given above, for the condition $\mathrm{D} = 0.353$ m this holds, as $C' \simeq 8.2$ bit/second for all 4 levels of ID. For the condition $\mathrm{D} = 0.212$ m, C' decreases when ID levels increase. Similarly, instructions to emphasize speed or accuracy should not significantly alter C'.

A two way ANOVA (Table 9.3) performed on the PD-dataset shows that indeed the effects of D and ID on the value of C' are non significant, hence results appear in line with the rationale that C' is not task related. Analysis of the G-dataset shows that the effect of instruction on the value of C' is also non significant at level $\alpha = 0.05$ ($F(4, 75) = 1.861$, $p = 0.126$, $\eta^2 = 0.09$). Hence, throughput as defined in (6.23) is independent of both task parameters and strategy, whereas, as remarked by Zhai [170], this is not the case for $\overline{\mathrm{TP}}$ 3.23, and therefore appears as a fundamental limit on our rate of transmission of information.

Table 9.3: Two-way ANOVA for effects of D and ID on C'.

	F	p	η^2
D	$F(1, 88) = 0.985$	0.324	0.011
ID	$F(3, 88) = 0.744$	0.529	0.024
DxID	$F(3, 88) = 0.779$	0.509	0.025

For completeness, we also looked at the effects of D, ID and instructions on the goodness of fit r^2 for the second phase. Analysis of the PD-dataset reveals a significant, yet small effect of ID (see Table 9.4) because higher values of ID lead to slightly better fits. The G-dataset reveals that instructions had a significant, moderate effect on r^2 ($F(4, 75) = 1.861$, $p < 0.001$, $\eta^2 = 0.352$). Thus, incentives to emphasize accuracy, be it through explicit instructions or task parameters, lead to slightly better empirical fits with the model.

A possible reason might be that higher accuracy requirements lead to longer durations for the second phase, as predicted by the model. Therefore, more points are considered in the regression, and the goodness of fit is mechanically increased.

Table 9.4: Two-way ANOVA for effects of D and ID on r^2.

	F	p	η^2
D	$F(1, 88) = 1.067$	0.305	0.011
ID	$F(3, 88) = 2.735$	0.048	0.084
DxID	$F(3, 88) = 0.232$	0.874	0.007

Final Matters

Conclusion

In spite of Fitts' law being a well established model for pointing in HCI, it was clear at the start of this work that there remained a lot of room in Fitts' law research to improve the theory (see Chapters 1 and 3). The main goal of this thesis was to propose a theoretical model for Fitts' law that would be more satisfying than the current state of the art's and provide novel tools for studying aiming movements in HCI. It was answered in two-steps.

1. We pushed the existing work *à la Fitts* as far as possible by proposing a formal information-theoretic model for Fitts' pointing task. Fitts' aiming task is modeled as a communication scheme, where each input sample represents one intention to be sent over the neuro-motor system. Starting from an error-less maximum entropy model (bounded uniform noise), we extended it to the case of pointing with errors with any noise distribution. This scheme rigorously defines a transmission channel whose capacity turns out to be exactly equal to the MacKenzie ID, thus legitimizing its use. In addition, that ID truly coincides with the celebrated Shannon capacity formula, which legitimizes the analogy with Shannon's Theorem 17. However, this model is completely black-box and does not account for feedback information.

2. In spite of our observation that human movement is incredibly diverse (high variability in all instances of movement, feedback and feedforward control, intermittent and continuous control), we have been capable of identifying an important invariant in the chronometric analysis of variance. The so-called PVP has been shown to be remarkably stereotyped in Fitts' paradigm, with a first phase of constant duration followed by a second phase, where the variance can decrease at best exponentially over time. This crucial result has been derived mathematically and validated on several empirical datasets.

These two theoretical models participate and should be further leveraged to provide a better understanding the issues discussed in Chapter 3:

- The link between the average and minimum time is explicit in the exGauss model. In particular, the whole scatter plot, including the front of performance and the average Fitts' law, is captured by β_0, β_1, k and σ^2.

- The effective model, which until now rested on a vague analogy with the entropy of a Gaussian, has been formalized in (6.29), without the "$+1$". This difference is only significant for very low values of ID.
- The FITTS 2 model provides a direct interpretation of the throughputs TP (3.22) and $\overline{\mathrm{TP}}$ (3.23). TP relates directly to the second phase of FITTS 2 where variance decreases, and characterize the part of aiming that ensures its reliability. $\overline{\mathrm{TP}}$ relates to the first and second phase of aiming, and is highly dependent on the paradigm effectively used and its implementation to measure pointing performance.

Beyond this purely theoretical work, two significant contributions ensue, that should prove useful to the practitioner as well as the researcher.

1. The **exGauss model**, which provides a model for endpoints, using 4 parameters that are easy to interpret, and which can fit adequately pointing data not only from controlled experiment but also from web-based, and even wild experiments. The exGauss model is expected to be a relevant tool for anyone who wishes to study pointing interactions outside of the laboratory.
2. The **FITTS 2 model**, which describes the whole trajectory, rather than endpoints only. This alone is a great improvement over current Fitts' law and opens many perspectives for pointing research. Additionally, the FITTS 2 model does not require a well defined target, as the PVP profile never assumes target width. Hence, FITTS 2 should as well prove useful for wild studies.

Some illustrations of the possibilities that emerge are given below.

Atypical Populations Studying how atypical populations produce aimed movements is paramount to improve the access of computing devices for special populations. Furthermore a stereotyped task such as Fitts' allows a precise characterization of pointing and could prove useful for a precise diagnosis of a pathology. Fig. 9.12 shows the PVPs for a typical adult (left panel), compared with two young dyspraxic users young wolf and young lemur (respectively middle and right panel). Young wolf's PVP is similar to the adult's one, with the exception that the maximum of standard deviation σ_O is about 100 times larger than the adult's. Young lemur's PVP not only has a similar incredibly high σ_0, but the decrease of variance is definitely not exponential, as the PVP is mostly convex. As a result, young lemur needs about twice as much time as the adult to finish his movement. The empirical pointing performance and PVP profiles might be explained by adapting the scheme in 6.3, e.g. by considering a noisy feedback link, or by evaluating the effect of removing a function in Fig. 6.5, e.g. the scaling α.

Input Techniques Analyzing pointing techniques with PVPs should prove insightful. Fig. 9.13 displays PVPs of several techniques using the BO dataset. The PVPs that are

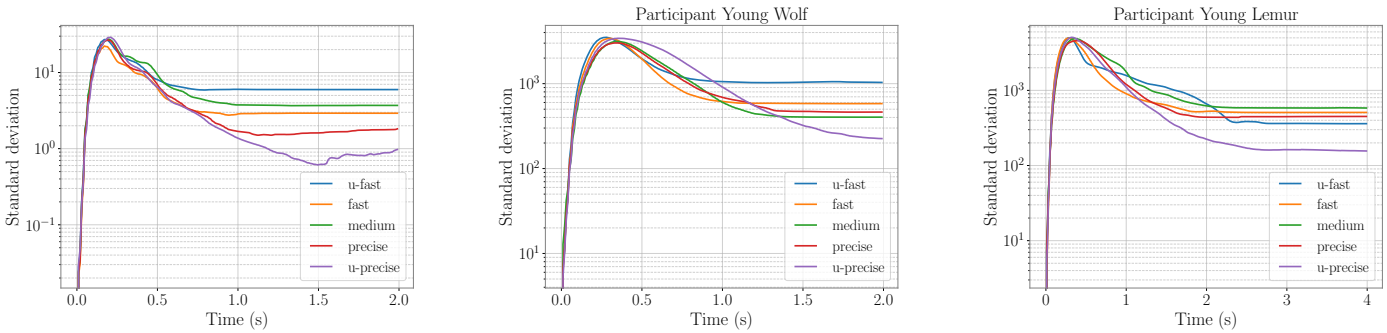

Figure 9.12: PVPs for 5 speed-accuracy conditions for one participant of the G-dataset and two dyspraxic children. Left: Typical adult user. Middle and Right: Two young dyspraxic users.

easy to interpret are those from the display space. This is expected, as aiming is based on visual feedback of the distance between the target and the cursor, hence the display space where the cursor lives is the one that matters. Note that this remark suggests that the trajectory in the motor space can be manipulated arbitrarily by adjusting the display of the cursor.

Observing the different PVPs reveal that the rake cursor completely skips the first phase of the PVP; in fact the techique inherently replaces the first phase by a pointing by the eye, which is much faster. However, notice that the gains are marginal in the end, as the first phase is short compared to the second phase, while the initial variance is higher than σ_0. Another observation that makes sense is that the Bubble cursor gains much of its speed as it reduces the needed accuracy at the end of the PVP; hence the second phase can be made shorter than for other techniques. A curious observation is that DynaSpot apparently perturbs the second phase, which is very non-linear. Semantic pointing is very close to Flat pointing. The difference between both techniques is that semantic pointing has an adjustable Control Display (CD) gain. Apparently, the only real gain this leads to is that a better accuracy can be reached in the end. What this suggests is that we (humans) easily adapt to changes in CD gain. I suspect that CD gain is somewhat analogous to stimulus response compatibility [41]: although there is an optimum CD gain function, with enough learning the effects of the CD function (within a certain range) can be made to disappear. In other words, the hypothesis is that the effects of the CD function are learned over time and the feedforward response is adapted appropriately.

As no target needs to be defined to computed PVPs, FITTS 2 is a good candidate to study “wild” aiming as well.

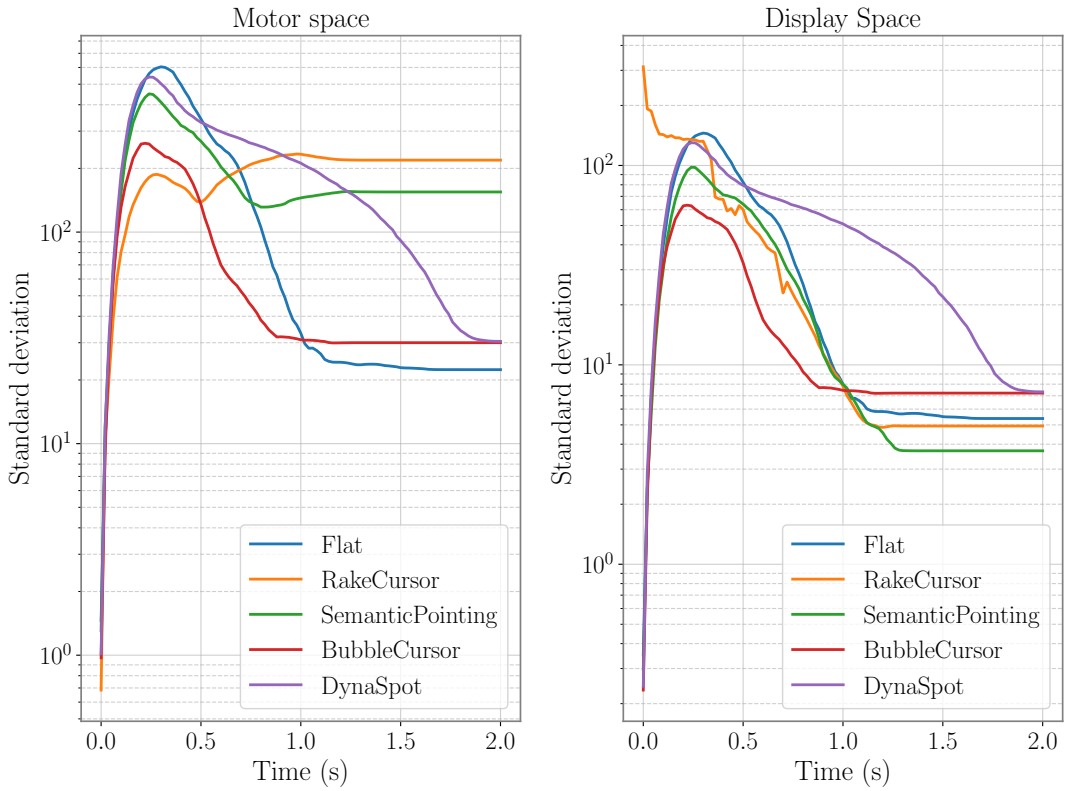

Figure 9.13: PVPs for 5 different techniques explored in the BO dataset. The left panel gives the PVPs for the motor space, while the right panel gives the PVPs for the display space.

List of contributions

Contributions are listed in chronological order.

1. European Mathematical Psychology Group (EMPG 2015): “Reconciling Fitts’ law with Shannon’s information theory” [60]
2. ACM Conference on Human Factors in Computing Systems (CHI ’17): “To Miss is Human: Information-Theoretic Rationale for Target Misses in Fitts’ Law” [55]
3. International Society for Pharmacoeconomics and Outcomes Research (ISPOR 2017): “Method Of Extrapolation: Estimation Of The Impact Of The Proportional Hazard Assumption On The Efficacy Estimation” [28] (participated throughout the project)
4. IFIP Conference on Human-Computer Interaction (Interact ’17): “One Fitts’ Law, Two Metrics” [57]

5. ACM Conference on Human Factors in Computing Systems (CHI '18): "The Perils of Confounding Factors: How Fitts' Law Experiments can Lead to False Conclusions" [58]

6. IEEE International Conference on Systems, Man, Cybernetics (SMC 2018): "Information-Theoretic Analysis of the Speed-Accuracy Tradeoff with Feedback" [59]

7. ACM Transactions on Computer Human Interaction (TOCHI): "Speed-Accuracy Tradeoff: A Formal Information-Theoretic Transmission Scheme (FITTS)" [56]

8. IEEE Transaction on Human Machine Systems (THMS): "A Feedback Information-Theoretic Transmission Scheme (FITTS) for Modelling Aimed Movements" [54] (submitted)

Bibliography

[1] 9241-9:2000, I. Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices. Standard, International Organization for Standardization, Geneva, Switzerland, Feb. 2000.

[2] AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike*. Springer, 1998, pp. 199–213.

[3] ATTNEAVE, F. *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* H. Holt, New York, 1959.

[4] BALAKRISHNAN, R. "beating" fitts' law: virtual enhancements for pointing facilitation. *International Journal of Human-Computer Studies 61*, 6 (2004), 857–874.

[5] BEGGS, W. D. A., AND HOWARTH, C. I. The accuracy of aiming at a target: Some further evidence for a theory of intermittent control. *Acta psychologica 36*, 3 (1972), 171–177.

[6] BLANCH, R., GUIARD, Y., AND BEAUDOUIN-LAFON, M. Semantic pointing: improving target acquisition with control-display ratio adaptation. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2004), ACM, pp. 519–526.

[7] BLANCH, R., AND ORTEGA, M. Rake cursor: improving pointing performance with concurrent input channels. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009), ACM, pp. 1415–1418.

[8] BULLOCK, D., AND GROSSBERG, S. Neural dynamics of planned arm movements: emergent invariants and speed-accuracy properties during trajectory formation. *Psychological review 95*, 1 (1988), 49.

[9] BURNHAM, K. P., AND ANDERSON, D. R. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media, 2003.

[10] CARD, S. K. *The psychology of human-computer interaction*. CRC Press, 2017.

[11] CARD, S. K., ENGLISH, W. K., AND BURR, B. J. Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a crt. *Ergonomics 21*, 8 (1978), 601–613.

[12] CARLTON, L. G. Processing visual feedback information for movement control. *Journal of Experimental Psychology: Human Perception and Performance 7*, 5 (1981), 1019.

[13] CARLTON, L. G. Visual processing time and the control of movement. *Advances in psychology 85* (1992), 3–31.

[14] CHAPUIS, O., BLANCH, R., AND BEAUDOUIN-LAFON, M. Fitts' Law in the Wild: A Field Study of Aimed Movements. article 1480, "LRI, December 2007. 11 pages.

[15] CHAPUIS, O., LABRUNE, J.-B., AND PIETRIGA, E. Dynaspot: speed-dependent area cursor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009), ACM, pp. 1391–1400.

[16] CHUA, R., AND ELLIOTT, D. Visual regulation of manual aiming. *Human Movement Science 12*, 4 (1993), 365 – 401.

[17] COCKBURN, A., GUTWIN, C., AND GREENBERG, S. A predictive model of menu performance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2007), CHI '07, ACM, pp. 627–636.

[18] COVER, T. M., AND THOMAS, J. A. *Elements of information theory*. John Wiley & Sons, 2012.

[19] CRONBACH, L. J. On the non-rational application of information measures in psychology. *Information Theory in Psychology Problems and Methods; Quastler, H., Ed* (1955), 14–30.

[20] CROSSMAN, E. R. F. W. *The Measurement of Perceptual Load in Manual Operations*. PhD thesis, University of Birmingham, 1956.

[21] CROSSMAN, E. R. F. W. The speed and accuracy of simple hand movements. Tech. rep., MRC and DSIR Joint Committee on Individual Efficiency in Industry, 1957. in *Nature and acquisition of industrial skills - DSIR 17/573*.

[22] CROSSMAN, E. R. F. W., AND GOODEVE, P. J. Feedback control of hand-movement and fitts' law. *The Quarterly Journal of Experimental Psychology 35*, 2 (1983), 251–278. First presented at the Oxford Meeting of the Experimental Psychology Society, July 1963.

[23] DARLING, W. G., AND COOKE, J. D. Changes in the variability of movement trajectories with practice. *Journal of Motor Behavior 19*, 3 (1987), 291–309.

[24] DARLING, W. G., AND COOKE, J. D. A linked muscular activation model for movement generation and control. *Journal of Motor Behavior 19*, 3 (1987), 333–354.

[25] DESMURGET, M., AND GRAFTON, S. Forward modeling allows feedback control for fast reaching movements. *Trends in cognitive sciences 4*, 11 (2000), 423–431.

[26] DESMURGET, M., TURNER, R. S., PRABLANC, C., RUSSO, G. S., ALEXANDER, G. E., AND GRAFTON, S. T. Updating target location at the end of an orienting saccade affects the characteristics of simple point-to-point movements. *Journal of Experimental Psychology: Human Perception and Performance 31*, 6 (2005), 1510.

[27] DREWES, H. Only one fitts' law formula please! In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (New York, NY, USA, 2010), CHI EA '10, ACM, pp. 2813–2822.

[28] DUTEIL, E., KANDEL, M., GORI, J., ROZE, S., AND BORGET, I. Method of extrapolation: Estimation of the impact of the proportional hazard assumption on the efficacy estimation. *Value in Health 20*, 9 (2017), A756–A757.

[29] ELIAS, P. Channel capacity without coding. In *Proceedings of the Institude of Radio Engineers* (1957), vol. 45, pp. 381–381.

[30] ELIAS, P. Two famous papers. *IRE Transactions on Information Theory 4*, 3 (1958), 99.

[31] ELKIND, J. I., AND FORGIE, C. D. Characteristics of the human operator in simple manual control systems. *IRE Transactions on Automatic Control*, 1 (1959), 44–55.

[32] ELLIOTT, D., CHUA, R., POLLOCK, B. J., AND LYONS, J. Optimizing the use of vision in manual aiming: The role of practice. *The Quarterly Journal of Experimental Psychology Section A 48*, 1 (1995), 72–83.

[33] ELLIOTT, D., HANSEN, S., GRIERSON, L. E. M., LYONS, J., BENNETT, S. J., AND HAYES, S. J. Goal-directed aiming: two components but multiple processes. *Psychological bulletin 136*, 6 (2010), 1023.

[34] ELLIOTT, D., HELSEN, W. F., AND CHUA, R. A century later: Woodworth's (1899) two-component model of goal-directed aiming. *Psychological bulletin 127*, 3 (2001), 342.

[35] ELLIOTT, D., LYONS, J., HAYES, S. J., BURKITT, J. J., ROBERTS, J. W., GRIERSON, L. E. M., HANSEN, S., AND BENNETT, S. J. The multiple process model of goal-directed reaching revisited. *Neuroscience & Biobehavioral Reviews 72* (2017), 95–110.

[36] FENN, W. O. The mechanics of muscular contraction in man. *Journal of Applied Physics 9*, 3 (1938), 165–177.

[37] FITTS, P. M. *The influence of response coding on performance in motor tasks, in Current Trends in Information Theory.* University of Pittsburgh Press, Pittsburgh, PA, 1953.

[38] FITTS, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology 47*, 6 (1954), 381.

[39] FITTS, P. M. Cognitive aspects of information processing: Iii. set for speed versus accuracy. *Journal of experimental psychology 71*, 6 (1966), 849.

[40] FITTS, P. M., AND PETERSON, J. R. Information capacity of discrete motor responses. *Journal of experimental psychology 67*, 2 (1964), 103.

[41] FITTS, P. M., AND POSNER, M. I. *Human performance.* Brooks/Cole, 1967.

[42] FITTS, P. M., AND RADFORD, B. K. Information capacity of discrete motor responses under different cognitive sets. *Journal of Experimental Psychology 71*, 4 (1966), 475.

[43] FLASH, T., AND HOGAN, N. The coordination of arm movements: an experimentally confirmed mathematical model. *Journal of neuroscience 5*, 7 (1985), 1688–1703.

[44] FRANKS, I. M., SANDERSON, D. J., AND VAN DONKELAAR, P. A comparison of directly recorded and derived acceleration data in movement control research. *Human Movement Science 9*, 6 (1990), 573 – 582.

[45] FREEDMAN, D. A. Ecological inference and the ecological fallacy. *International Encyclopedia of the social & Behavioral sciences 6*, 4027-4030 (1999), 1–7.

[46] GABOR, D. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers - Part III: Radio and Communication Engineering 93*, 26 (November 1946), 429–441.

[47] GALLAGER, R. G., AND NAKIBOGLU, B. Variations on a theme by schalkwijk and kailath. *IEEE Transactions on Information Theory 56*, 1 (2010), 6–17.

[48] GAN, K., AND HOFFMANN, E. R. Geometrical conditions for ballistic and visually controlled movements. *Ergonomics 31*, 5 (1988), 829–839.

[49] GILBERT, E. N. Capacity of a burst-noise channel. *Bell Sys. Tech. J. 39* (1960), 1253–1265.

[50] GILLAN, D. J., HOLDEN, K., ADAM, S., RUDISILL, M., AND MAGEE, L. How should fitts' law be applied to human-computer interaction? *Interacting with Computers 4*, 3 (1992), 291–313.

[51] GOGUEY, A., CASIEZ, G., COCKBURN, A., AND GUTWIN, C. Storyboard-based empirical modeling of touch interface performance. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 445.

[52] GOLDBERG, K., FARIDANI, S., AND ALTEROVITZ, R. Two large open-access datasets for fitts' law of human motion and a succinct derivation of the square-root variant. *IEEE Transactions on Human-Machine Systems 45*, 1 (2015), 62–73.

[53] GORDON, J., AND GHEZ, C. Trajectory control in targeted force impulses. *Experimental brain research 67*, 2 (1987), 253–269.

[54] GORI, J., AND RIOUL, O. A feedback information-theoretic transmission scheme (FITTS) for modelling aimed movements. *CoRR abs/1804.05021* (2018).

[55] GORI, J., RIOUL, O., AND GUIARD, Y. To miss is human: Information-theoretic rationale for target misses in fitts' law. In *CHI'17: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Denver, USA, May 2017), ACM, p. 5.

[56] GORI, J., RIOUL, O., AND GUIARD, Y. Speed-accuracy tradeoff: A formal information-theoretic transmission scheme (fitts). *ACM Trans. Comput.-Hum. Interact. 25*, 5 (Sept. 2018), 27:1–27:33.

[57] GORI, J., RIOUL, O., GUIARD, Y., AND BEAUDOUIN-LAFON, M. One fitts' law, two metrics. In *IFIP Conference on Human-Computer Interaction* (Heidelberg, Germany, 2017), Springer, pp. 525–533.

[58] GORI, J., RIOUL, O., GUIARD, Y., AND BEAUDOUIN-LAFON, M. The perils of confounding factors: How fitts' law experiments can lead to false conclusions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018), ACM, p. 196.

[59] GORI, J. AND RIOUL, O. Information-Theoretic Analysis of the Speed-Accuracy Tradeoff with Feedback. In *Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC 2018)* (2018), IEEE.

[60] GORI, J. AND RIOUL, O. AND GUIARD, Y. Reconciling Fitts' law with Shannon's information theory. European Mathematical Psychology Group Meeting (EMPG 2015). Padua, Italy, Sept. 1-3, 2015.

[61] GREEN, D. M., AND SWETS, J. A. *Signal Detection Theory and Psychophysics*. John Wiley & Sons, 1989.

[62] GROSS, J., TIMMERMANN, L., KUJALA, J., DIRKS, M., SCHMITZ, F., SALMELIN, R., AND SCHNITZLER, A. The neural basis of intermittent motor control in humans. *Proceedings of the National Academy of Sciences 99*, 4 (2002), 2299–2302.

[63] GROSSMAN, T., AND BALAKRISHNAN, R. The bubble cursor: enhancing target acquisition by dynamic resizing of the cursor's activation area. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2005), ACM, pp. 281–290.

[64] GUIARD, Y., AND BEAUDOUIN-LAFON, M., Eds. *Fitts' law 50 years later: applications and contributions from human-computer interaction* (2004), vol. 61.

[65] GUIARD, Y., AND OLAFSDOTTIR, H. B. On the measurement of movement difficulty in the standard approach to fitts' law. *PLoS one 6*, 10 (2011), e24389.

[66] GUIARD, Y., OLAFSDOTTIR, H. B., AND PERRAULT, S. T. Fitt's law as an explicit time/error trade-off. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, 2011), ACM, pp. 1619–1628.

[67] GUIARD, Y., AND RIOUL, O. A mathematical description of the speed/accuracy trade-off of aimed movement. In *Proceedings of the 2015 British HCI Conference* (New York, 2015), ACM, pp. 91–100.

[68] GUPTA, K. L. Aggregation bias in linear economic models. *International Economic Review* (1971), 293–305.

[69] GUTMAN, S. R., AND GOTTLIEB, G. L. Basic functions of variability of simple pre-planned movements. *Biological Cybernetics 68*, 1 (1992), 63–73.

[70] HAKE, H. W. A note on the concept of "channel capacity" in psychology. In *Information Theory in Psychology: Problems and Methods. Proceedings of a Conference on the Estimation of Information Flow, Monticello, Illinois, July 5-9, 1954, and Related Papers* (1955), Free Press, p. 248.

[71] HANCOCK, P. A., AND NEWELL, K. M. The movement speed-accuracy relationship in space-time. *Motor Behaviour: Programming, Control, and cquisition* (1985), 153–188.

[72] HARRIS, C. M., AND WOLPERT, D. M. Signal-dependent noise determines motor planning. *Nature 394*, 6695 (1998), 780.

[73] HARTIGAN, J. A., AND HARTIGAN, P. M. The dip test of unimodality. *The annals of Statistics 13*, 1 (1985), 70–84.

[74] HICK, W. E. On the rate of gain of information. *Quarterly Journal of experimental psychology 4*, 1 (1952), 11–26.

[75] HOFFMANN, E. R. Which version/variation of Fitts' law? a critique of information-theory models. *Journal of motor behavior 45*, 3 (2013), 205–215.

[76] HURST, A., HUDSON, S. E., AND MANKOFF, J. Dynamic detection of novice vs. skilled use without a task model. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2007), CHI '07, ACM, pp. 271–280.

[77] HYMAN, R. Stimulus information as a determinant of reaction time. *Journal of experimental psychology 45*, 3 (1953), 188.

[78] JAGACINSKI, R. J., REPPERGER, D. W., MORAN, M. S., WARD, S. L., AND GLASS, B. Fitts' law and the microstructure of rapid discrete movements. *Journal of Experimental Psychology: Human Perception and Performance 6*, 2 (1980), 309.

[79] JERISON, D., SINGER, L. M., AND STROOCK, D. W. The legacy of norbert weiner: A centennial symposium in honour of the 100th anniversary of norbert weiner's birth. In *Proceedings of Symposia in Pure Mathematics* (1994), vol. 60.

[80] JOTA, R., NG, A., DIETZ, P., AND WIGDOR, D. How fast is fast enough?: a study of the effects of latency in direct-touch pointing tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), ACM, pp. 2291–2300.

[81] JUDE, A., GUINNESS, D., AND POOR, G. M. Reporting and visualizing fitts's law: Dataset, tools and methodologies. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (2016), ACM, pp. 2519–2525.

[82] KANDEL, E. R., SCHWARTZ, J. H., JESSELL, T. M., SIEGELBAUM, S., AND HUDSPETH, A. J. *Principles of neural science*, vol. 4. McGraw-hill New York, 2000.

[83] KEELE, S. W. Movement control in skilled motor performance. *Psychological bulletin 70*, 6p1 (1968), 387.

[84] KHAN, M. A., FRANKS, I. M., ELLIOTT, D., LAWRENCE, G. P., CHUA, R., BERNIER, P., HANSEN, S., AND WEEKS, D. J. Inferring online and offline processing of visual feedback in target-directed movements from kinematic data. *Neuroscience & Biobehavioral Reviews 30*, 8 (2006), 1106–1121.

[85] KIM, M., KIM, B. H., AND JO, S. Quantitative evaluation of a low-cost noninvasive hybrid interface based on eeg and eye movement. *IEEE transactions on neural systems and rehabilitation engineering 23*, 2 (2015), 159–168.

[86] KVÅLSETH, T. O. An alternative to fitts' law. *Bulletin of the Psychonomic Society 16*, 5 (Nov 1980), 371–373.

[87] LAI, S., MAYER-KRESS, G., SOSNOFF, J. J., AND NEWELL, K. M. Information entropy analysis of discrete aiming movements. *Acta Psychologica 119*, 3 (2005), 283–304.

[88] LAMING, D. Statistical information and uncertainty: A critique of applications in experimental psychology. *Entropy 12*, 4 (2010), 720–771.

[89] LANGOLF, G. D., CHAFFIN, D. B., AND FOULKE, J. A. An investigation of fitts' law using a wide range of movement amplitudes. *Journal of Motor Behavior 8*, 2 (1976), 113–128.

[90] LEVINE, R. D., AND TRIBUS, M. Maximum entropy formalism. In *Maximum Entropy Formalism Conference (1978: Massachusetts Institute of Technology)* (1979), Mit Press.

[91] LUCE, R. D. The theory of selective information and some of its behavioral applications. *Developments in mathematical psychology* (1960), 5–119.

[92] LUCE, R. D. Whatever happened to information theory in psychology? *Review of General Psychology 7*, 2 (2003), 183–188.

[93] MACKENZIE, I. S. A note on the information-theoretic basis for fitts' law. *Journal of motor behavior 21*, 3 (1989), 323–330.

[94] MACKENZIE, I. S. *Fitts' law as a performance model in human-computer interaction*. PhD thesis, University of Toronto, 1992.

[95] MACKENZIE, I. S. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction 7*, 1 (1992), 91–139.

[96] MACKENZIE, I. S. A note on the validity of the shannon formulation for fitts' index of difficulty. *Open Journal of Applied Sciences 3*, 06 (2013), 360.

[97] MACKENZIE, I. S., AND BUXTON, W. Extending fitts' law to two-dimensional tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (New York, 1992), ACM, pp. 219–226.

[98] MACKENZIE, I. S., AND JUSOH, S. An evaluation of two input devices for remote pointing. In *Engineering for human-computer interaction*. Springer, 2001, pp. 235–250.

[99] MCINTOSH, R. D., MON-WILLIAMS, M., AND TRESILIAN, J. R. Grasping at laws: Speed-accuracy trade-offs in manual prehension. *Journal of experimental psychology: human perception and performance 44*, 7 (2017), 1022.

[100] MCRUER, D. Human dynamics in man-machine systems. *Automatica 16*, 3 (1980), 237–253.

[101] MESSIER, J., AND KALASKA, J. F. Comparison of variability of initial kinematics and endpoints of reaching movements. *Experimental Brain Research 125*, 2 (1999), 139–152.

[102] MEYER, D. E., ABRAMS, R. A., KORNBLUM, S., WRIGHT, C. E., AND KEITH SMITH, J. E. Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological review 95*, 3 (1988), 340.

[103] MEYER, D. E., KEITH-SMITH, J. E., KORNBLUM, S., ABRAMS, R. A., AND WRIGHT, C. E. *Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action.* Lawrence Erlbaum Associates, Inc, 1990.

[104] MILLER, G. A. Communication. *Annual Review of Psychology*, 5 (1954), 401–420. Stone, C. P. and McNemar, Q., eds.

[105] MILLER, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review 63*, 2 (1956), 81.

[106] MILLER, G. A., AND FRICK, F. C. Statistical behavioristics and sequences of responses. *Psychological Review 56*, 6 (1949), 311.

[107] MOTTET, D., VAN DOKKUM, L. E. H., FROGER, J., GOUAÏCH, A., AND LAFFONT, I. Trajectory formation principles are the same after mild or moderate stroke. *PloS one 12*, 3 (2017), e0173674.

[108] MOULTON, H. G. Businessman's briefs. *The Satturday Review* (1950).

[109] MÜLLER, J., OULASVIRTA, A., AND MURRAY-SMITH, R. Control theoretic models of pointing. *ACM Transactions on Computer-Human Interaction (TOCHI) 24*, 4 (2017), 27.

[110] NELSON, W. L. Physical principles for economies of skilled movements. *Biological cybernetics 46*, 2 (1983), 135–147.

[111] NEVIN, J. A. Signal detection theory and operant behavior: A review of david m. green and john a. swets' signal detection theory and psychophysics. 1. *Journal of the Experimental Analysis of Behavior 12*, 3 (1969), 475–480.

[112] NEWELL, A., AND CARD, S. K. The prospects for psychological science in human-computer interaction. *Human-computer interaction 1*, 3 (1985), 209–242.

[113] OLAFSDOTTIR, H. B., GUIARD, Y., RIOUL, O., AND PERRAULT, S. T. A new test of throughput invariance in fitts' law: role of the intercept and of jensen's inequality. In *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers* (2012), British Computer Society, pp. 119–126.

[114] PARTRIDGE, L. D. Signal-handling characteristics of load-moving skeletal muscle. *American Journal of Physiology-Legacy Content 210*, 5 (1966), 1178–1191.

[115] PÉLISSON, D., PRABLANC, C., GOODALE, M. A., AND JEANNEROD, M. Visual control of reaching movements without vision of the limb. *Experimental Brain Research 62*, 2 (1986), 303–311.

[116] PINKER, S. George a. miller (1920–2012)., 2013. Obituary.

[117] PLAMONDON, R., AND ALIMI, A. M. Speed/accuracy trade-offs in target-directed movements. *Behavioral and brain sciences 20*, 2 (1997), 279–303.

[118] POLIT, A., AND BIZZI, E. Processes controlling arm movements in monkeys. *Science 201*, 4362 (1978), 1235–1237.

[119] POPPER, K. *Realism and the aim of science: From the postscript to the logic of scientific discovery*. Routledge, 2013.

[120] PRABLANC, C., ECHALLIER, J. E., JEANNEROD, M., AND KOMILIS, E. Optimal response of eye and hand motor systems in pointing at a visual target. ii. static and dynamic visual cues in the control of hand movement. *Biological cybernetics 35*, 3 (1979), 183–187.

[121] PROTEAU, L., MARTENIUK, R. G., GIROUARD, Y., AND DUGAS, C. On the type of information used to control and learn an aiming movement after moderate and extensive training. *Human Movement Science 6*, 2 (1987), 181–199.

[122] QUASTLER, H. Approximate estimation of information measures. In *Information Theory in Psychology: Problems and Methods. Proceedings of a Conference on the Estimation of Information Flow, Monticello, Illinois, July 5-9, 1954, and Related Papers* (1955), Free Press, p. 124.

[123] RÉDEI, M. *John von Neumann: selected letters*, vol. 27. American Mathematical Soc., 2005.

[124] RIOUL, O., AND GUIARD, Y. The power model of Fitts' law does not encompass the logarithmic model. *Electronic Notes in Discrete Mathematics 42* (2013), 65–72.

[125] RIOUL, O., AND MAGOSSI, J. C. On Shannon's formula and Hartley's rule: Beyond the mathematical coincidence. *Entropy 16*, 9 (2014), 4892–4910.

[126] ROBERTS, S., AND PASHLER, H. How persuasive is a good fit? a comment on theory testing. *Psychological review 107*, 2 (2000), 358.

[127] ROSENBAUM, D. A. *Human motor control*. Academic press, 2009.

[128] SAUNDERS, J. A., AND KNILL, D. C. Humans use continuous visual feedback from the hand to control fast reaching movements. *Experimental brain research 152*, 3 (2003), 341–352.

[129] SCHMIDT, R. A., ZELAZNIK, H., HAWKINS, B., FRANK, J. S., AND QUINN JR, J. T. Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological review 86*, 5 (1979), 415.

[130] SEGAL, J. *Le Zéro et le Un : Histoire de la notion scientifique d'information au 20e siècle*. Editions Syllepse, 2003.

[131] SEOW, S. C. Information theoretic models of hci: a comparison of the hick-hyman law and fitts' law. *Human-Computer Interaction 20*, 3 (2005), 315–352.

[132] SHANNON, C. E. A mathematical theory of communication, part i, part ii. *Bell Syst. Tech. J. 27* (1948), 623–656.

[133] SHANNON, C. E. The bandwagon. *IRE Transactions on Information Theory 2*, 1 (1956), 3.

[134] SHANNON, C. E., AND WEAVER, W. *The mathematical theory of communication (1949).* Urbana, IL: University of Illinois Press, 1963.

[135] SHANNON, C. E., AND WEAVER, W. *La théorie Mathématique de la Communication*. Cassini, 2018. (French Translation by Cosnier, J. and Dahan, G. and Economidès, S. and Bellaïche, C. and Rioul, O.).

[136] SHERIDAN, M. R. A reappraisal of fitts' law. *Journal of Motor Behavior 11*, 3 (1979), 179–188.

[137] SHMUELOF, L., KRAKAUER, J. W., AND MAZZONI, P. How is a motor skill learned? change and invariance at the levels of task success and trajectory control. *Journal of neurophysiology 108*, 2 (2012), 578–594.

[138] SIGCHI. 2001 SIGCHI Awards (incomplete). `https://web.archive.org/web/20150907183121/http://www.sigchi.org/about/awards/awards-2001.html`. Accessed: 15-10-2018.

[139] SMITH, J. E. K. Rapid aimed movements and the speed-accuracy trade off: Optimal strategies, 1988. C. Coombs Memorial Symp., in Brown, D. R. and Smith, J. E. K. (Eds.), Frontiers of mathematical psychology: Essays in honor of Clyde Coombs.

[140] SOECHTING, J. F. Effect of target size on spatial and temporal characteristics of a pointing movement in man. *Experimental Brain Research 54*, 1 (1984), 121–132.

[141] SOECHTING, J. F., AND LACQUANITI, F. Invariant characteristics of a pointing movement in man. *Journal of Neuroscience 1*, 7 (1981), 710–720.

[142] SOUKOREFF, R. W., AND MACKENZIE, I. S. Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci. *International journal of human-computer studies 61*, 6 (2004), 751–789.

[143] SOUKOREFF, R. W., ZHAO, J., AND REN, X. The entropy of a rapid aimed movement: Fitts' index of difficulty versus shannon's entropy. In *IFIP Conference on Human-Computer Interaction* (Heidelberg, Germany, 2011), Springer, pp. 222–239.

[144] STUBBS, D. F. What the eye tells the hand. *Journal of Motor Behavior 8*, 1 (1976), 43–58.

[145] THRUSTON, J. B. Devaluing the human brain. *The Satturday Review* (1949).

[146] TODOROV, E., AND JORDAN, M. I. Optimal feedback control as a theory of motor coordination. *Nature neuroscience 5*, 11 (2002), 1226–1235.

[147] TODOROV, E. V. *Studies of goal directed movements*. PhD thesis, Massachusetts Institute of Technology, 1998.

[148] VAN BEERS, R. J., HAGGARD, P., AND WOLPERT, D. M. The role of execution noise in movement variability. *Journal of neurophysiology 91*, 2 (2004), 1050–1063.

[149] VAN DER MEULEN, J. H. P., GOOSKENS, R. H. J. M., DENIER VAN DER GON, J. J., GIELEN, C. C. A. M., AND WILHELM, K. Mechanisms underlying accuracy in fast goal-directed arm movements in man. *Journal of Motor Behavior 22*, 1 (1990), 67–84.

[150] VAN GALEN, G. P., AND DE JONG, W. P. Fitts' law as the outcome of a dynamic noise filtering model of motor control. *Human movement science 14*, 4 (1995), 539–571.

[151] VAN GALEN, G. P., AND SCHOMAKER, L. R. B. Fitts' law as a low-pass filter effect of muscle stiffness. *Human movement science 11*, 1 (1992), 11–21.

[152] VAN GALEN, G. P., VAN DOORN, R. R., AND SCHOMAKER, L. R. Effects of motor programming on the power spectral density function of finger and wrist movements. *Journal of Experimental Psychology: Human Perception and Performance 16*, 4 (1990), 755.

[153] VAN HUFFEL, S., AND LEMMERLING, P. *Total least squares and errors-in-variables modeling: analysis, algorithms and applications*. Springer Science & Business Media, 2013.

[154] WAGENMAKERS, E., AND BROWN, S. On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological review 114*, 3 (2007), 830.

[155] WALLACE, S. A., AND NEWELL, K. M. Visual control of discrete aiming movements. *The Quarterly Journal of Experimental Psychology 35*, 2 (1983), 311–321.

[156] WARTON, D. I., WRIGHT, I. J., FALSTER, D. S., AND WESTOBY, M. Bivariate line-fitting methods for allometry. *Biological Reviews 81*, 2 (2006), 259–291.

[157] WELFORD, A. T. The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics 3*, 3 (1960), 189–230.

[158] WELFORD, A. T. *Fundamentals of skill.* Methuen, London, 1968.

[159] WELFORD, A. T., NORRIS, A. H., AND SHOCK, N. W. Speed and accuracy of movement and their changes with age. *Acta psychologica 30* (1969), 3–15.

[160] WICKELGREN, W. A. Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica 41*, 1 (1977), 67–85.

[161] WIENER, N. *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press, 1948. 2nd revised ed. 1961.

[162] WIENER, N. *I am a Mathematician The Later Life of a Prodigy*. The M.I.T. Press, 1956.

[163] WIENER, N. *The human use of human beings: Cybernetics and society*. Perseus Books Group, 1988.

[164] WIESENDANGER, M. Organization of secondary motor areas of cerebral cortex. *Handbook of physiology 2*, Part 2 (1981), 1121–1147.

[165] WOBBROCK, J. O., AND KIENTZ, J. A. Research contributions in human-computer interaction. *Interactions 23*, 3 (Apr. 2016), 38–44.

[166] WOBBROCK, J. O., MYERS, B. A., AND AUNG, H. H. The performance of hand postures in front-and back-of-device interaction for mobile computing. *International Journal of Human-Computer Studies 66*, 12 (2008), 857–875.

[167] WOODWORTH, R. S. Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements 3*, 3 (1899), i.

[168] ZELAZNIK, H. N., HAWKINS, B., AND KISSELBURGH, L. Rapid visual feedback processing in single-aiming movements. *Journal of motor behavior 15*, 3 (1983), 217–236.

[169] ZELAZNIK, H. N., SCHMIDT, R. A., AND GIELEN, S. C. A. M. Kinematic properties of rapid aimed hand movements. *Journal of Motor Behavior 18*, 4 (1986), 353–372.

[170] ZHAI, S. Characterizing computer input with fitts' law parameters—the information and non-information aspects of pointing. *International Journal of Human-Computer Studies 61*, 6 (2004), 791–809.

[171] ZHAI, S., KONG, J., AND REN, X. Speed–accuracy tradeoff in fitts' law tasks—on the equivalency of actual and nominal pointing precision. *International journal of human-computer studies 61*, 6 (2004), 823–856.

[172] ZHAI, S., KRISTENSSON, P. O., APPERT, C., ANDERSON, T. H., CAO, X., ET AL. Foundational issues in touch-surface stroke gesture design—an integrative review. *Foundations and Trends® in Human–Computer Interaction 5*, 2 (2012), 97–205.