Information-Theoretic Analysis of the Speed-Accuracy Tradeoff with Feedback

Julien Gori and Olivier Rioul
LTCI, Télécom ParisTech, Université Paris-Saclay
F-75013 Paris, France
firstname.lastname@telecom-paristech.fr

***Abstract*—Human movements are inherently variable and involve some feedback mechanism. A study of the positional variance in a tapping task reveals that the variance profiles are unimodal in time. In the variance-decreasing phase, the aiming task can be modeled by a Shannon-like communication scheme where information is transmitted from a "source"—determined by the distance to the target at maximum variance—to a "destination"—the movement endpoint—over a "channel" with feedback perturbed by Gaussian noise. Thanks to the feedback link, we show that the variance decreases exponentially at a rate given by the channel capacity. This is confirmed on real data. The proposed information-theoretic model has promise to improve our understanding of human aimed movements.**

I. Introduction

Variability is an important characteristic of human aimed movement. In fact, it has long been observed [1] that a participant will produce disparate movements when asked to repeatedly perform the same task, while having to slow down to perform more accurate tasks—the so-called *speed-accuracy tradeoff*. Many attempts at modeling variability were proposed in the literature, e.g., [2], [3] which only characterize the variability of the movement's *endpoints* based on Fitts' paradigm [4]. We focus instead on the variability of the position of the limb extremity used for aiming during the entire movement: How does the positional variance evolve over time?

As remarked in [5], trajectory variability has "surprisingly received little attention from researchers". The entropy of a set of trajectories over time was considered in [6], where unimodal profiles were observed. Positional variance was studied in [7], [8] but only at specific kinematic markers (peak acceleration, velocity, deceleration, and movement time). Complete variance profiles for elbow flexion were reported in [9], [10], most resulting variance profiles being unimodal—yet human movements seem much more complex as they involve many joints.

The following simple reasoning suggests that positional variance profiles should be *unimodal*. Consider e.g., Fitts' tapping task [4] where the participant aims at a target of a given width at a given distance:

1) all movements start at the same position: initial positional variance is null;
2) in the early stages of the movement, positional variance increases [9], [11];
3) if time permits, the participant eventually reaches the target [9], [12], [13] and the movement ends, which implies that the positional variance vanishes asymptotically [14].

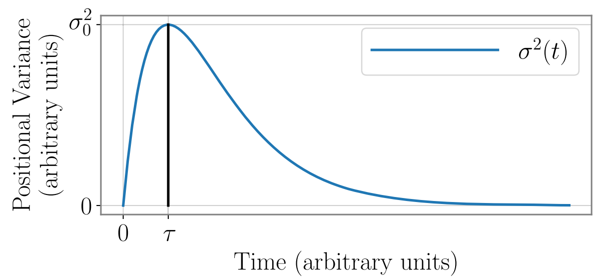

Fig. 1. Ideal positional variance profile with maximum variance σ_0^2 at $t = \tau$.

We therefore expect a *unimodal* variance profile as illustrated in Fig. 1: A first variance-increasing phase for time $\in [0, \tau]$, where variance increases from 0 to σ_0^2, is followed by a second variance-decreasing phase where the variance is decreased to arbitrarily low values and eventually vanishes.

Yet another important characteristic of human aiming is the use of feedback [15, p. 95] for generating movements. Feedback can be visual [16]–[18] or kinesthetic [19], [20] and affects movement as short as 100 ms [17], [20], [21]. Feedback is often associated to control theory, which has been applied to human aiming movements but cannot easily handle trajectory variability nor noise [22]. Now, feedback is also frequently used in Shannon's information theory [23], [24] to enhance communication in the presence of noise. Unfortunately, even though many previous studies have used information-theoretic tools to model human movements [4], [25]–[27], they did not account for feedback mechanisms. There wasn't even any explicit channel model described in Fitts' seminal paper [4], and most recent attempts like [26], [27] build on a feedforward channel without feedback. This may explain why information theory was said to be "no longer much of a factor" [28] for explaining human movements or more generally in experimental psychology.

In this paper, we present a complete information-theoretic model that uses feedback to predict that for $t > \tau$ (in the second variance-decreasing phase) the variability decreases exponentially at a rate given by Shannon's channel capacity. Our contributions are as follows. 1) Validate the unimodality hypothesis on empirical pointing data 2) Propose an information-theoretic model for the variance decreasing phase 3) Determine Shannon's capacity for this model, to provide an upper bound on the rate at which variance decreases over time 4) Validate the model and estimate the capacity on empirical pointing data for various participants.

2577-1655/18/$31.00 ©2018 IEEE
DOI 10.1109/SMC.2018.00585

II. Unimodal Variance Profiles in Real Data

In this section, we evaluate the unimodality hypothesis on empirical data. Since there have been numerous published studies on pointing tasks, we have simply re-analyzed an existing dataset (hereafter named "G-dataset") from a study by Guiard et al. [29] which uses a discrete version of Fitts' tapping task [30]. Using a stylus on a digital tablet, participants were instructed to reach a line located 150mm away from a starting point, under conditions ranging from (1) full speed emphasis to (5) full accuracy emphasis. The raw trajectory data was low-pass filtered and re-sampled closest to the average sampling frequency, and individual movements were extracted and synchronized so as to have a common time origin. 16 participants produced movements for five conditions, repeated five times, providing 80 different variance profiles[1]. As an illustration, Fig. 2 displays the trajectories of participant 3[2] of the dataset, performing under condition (3) balanced speed/accuracy.

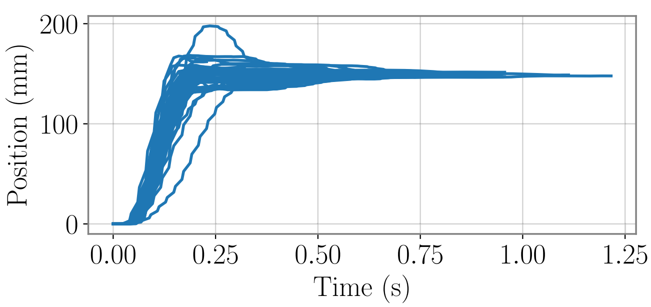

Fig. 2. The set of trajectories for participant 3 performing under condition 3 (balance speed versus accuracy) for the G-dataset.

The unimodality hypothesis was verified by computing the positional variance profiles for each subset composed of *all* the trajectories acquired by a given participant under the same condition[3]. We checked unimodality by ensuring that there is only one sign change in the time derivative of the profile. We found that 64 out of 80 variance profiles in the dataset are strictly unimodal. Fig. 3 displays a typical empirical variance profile consistent with the ideal profile given in Fig. 1, as well as its derivative in orange, obtained from the same set as the one used for Fig. 2. There are a few profiles with multiple modes that are often due to outliers: Usually, the secondary modes are very small and the variance profile can still be reasonably well modeled by a unimodal profile.

For completeness, we also conducted the Hartigan & Hartigan DIP test for unimodality [31], where the null hypothesis is that the distribution is unimodal. Only 9 profiles ($p<.05$) were found to be significantly non-unimodal out of the total 80 profiles and the

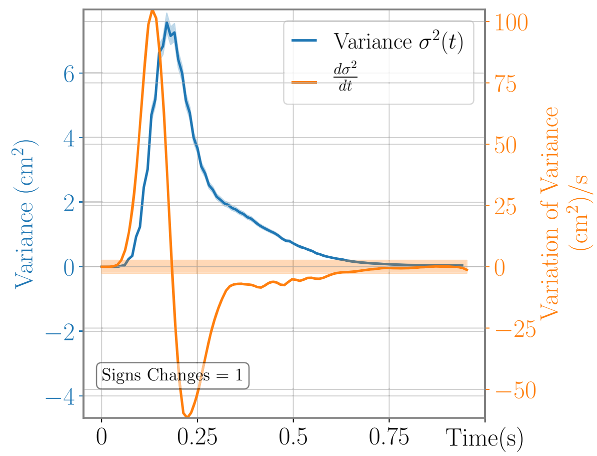

Fig. 3. The unimodal empirical variance profile (in blue) of participant 3 performing in the balanced condition (3). The mode appears for $\tau \simeq 200$ ms. The derivative of the variance profile (in orange) changes sign only once.

average p-value is just below 0.6. We also proceeded to a visual check and determined that 4 profiles that were labeled as multi-modal had a massive outlier which distorted the variance profile.

There is thus substantial evidence to support the unimodality of variance profiles in tapping tasks.

III. A Model for the Variance-Decreasing Phase

A. Information-Theoretic Model Description

Whereas most existing models aim at predicting an average trajectory, the goal here is to predict how the variability of a set of trajectories evolves over time. Asymptotically, i.e., when the set is large enough, the position of the limb extremity (the 'limb' in short) is a continuous random variable. At the end of the first phase, we may assume that the position can be well approximated by a Gaussian distribution[4], with some standard deviation σ_0 (see Fig. 1).

Thanks to the feedback, 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 [20]. Hence the position of the target is also known at the brain level. The distance from limb to target can thus easily evaluated by the brain; in fact it can be readily estimated by the eye if the limb is close enough to the target.

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. Moreover, for rapid aiming, the first phase may equally well undershoot or overshoot the target [34], and we may assume zero mean: $A \sim \mathcal{N}(0, \sigma_0^2)$. To successfully complete the movement, the brain has to send A to the limb; this is precisely the role of the second phase.

[1] No outliers were removed because removal procedures are based on arbitrary heuristics for which we found no satisfying method. As seen below, we nonetheless found compelling evidence in favor of the unimodality hypothesis.

[2] Throughout the paper we will be using the example of participant 3. The study was carried out for every participant in the G-dataset and all participants behaved similarly.

[3] As all traces end at different dates, the number of trajectories available upon which variance can be computed decreases with time; to ensure reliable estimations, we stopped computing variance when less than 10 trajectories were available.

[4] This is an assumption that is often used [11], [32], whose strict veracity depends on the conditions in which the movements were produced [33]. Usually, this approximation is good enough as distributions are bell-shaped. In [6], the empirical entropy of the trajectory was compared with the theoretical entropy of a Gaussian distribution. The difference was never more than .3 bits, throughout the entire trajectory.

We consider the following scheme:

- From A the brain outputs a certain amplitude X_1 to be sent to the limb:
$$X_1 = f(A), \tag{1}$$
where f is an unknown function performed by the brain.
- To account for the variability of the human motor system [35], we consider a noisy transmission from brain to limb, where X_1 gets perturbed by additive white Gaussian noise (so-called AWGN channel). The output of the channel Y_1 is given by
$$Y_1 = X_1 + Z_1 \quad \text{where } Z_1 \sim \mathcal{N}(0,N). \tag{2}$$
- Based on the channel output Y_1, distance $\widehat{A}_1$ is actually covered, which is the result of some unknown 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. Each such step is infinitesimal and thus, the whole process is an intermittent iterative correction model that becomes continuous at the limit.

At iteration i, the scheme is described by following equations (see Fig. 4):

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 [32], or to the second component in two-component models [1], [19], [20].

B. Bounds on Transmitted Information

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 Shannon capacity C [23], [24] of the AWGN Channel under power constraint $P_i \leq P$ and noise power N is given by
$$C = \frac{1}{2}\log_2(1 + P/N). \tag{3}$$
expressed in bits per channel use.
- $D_n = \mathbb{E}[(A - \widehat{A}_n)^2]$ is the quadratic *distortion* that represents the mean-squared error of the estimation of A by $\widehat{A}_n$ after n iterations (channel uses); the distortion essentially corresponds to the empirical variance.
- $I(A, \widehat{A}_n)$ is Shannon's mutual information [24] between A and $\widehat{A}_n$.

Theorem 1. *Consider the transmission scheme of Fig. 4 with an AWGN channel of Shannon capacity C and noiseless feedback. For a zero-mean Gaussian source A with variance σ_0^2, we have that:*
$$\frac{1}{2}\log\frac{\sigma_0^2}{D_n} \underset{(a)}{\leq} I(A, \widehat{A}_n) \underset{(b)}{\leq} nC. \tag{4}$$

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.

IV. Achieving Capacity

For a given channel C and number of channel uses n, 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 (4):
$$\frac{1}{2}\log\frac{\sigma_0^2}{D_n} = I(M, \widehat{M}) = nC. \tag{5}$$

The goal of this section is to find the scheme that achieves optimality (Eq. (5)).

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

Lemma 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 are independent from the errors $A - \widehat{A}^i$;*
5) *$\widehat{A}_i = g(Y^i)$ is a sufficient statistic of Y^i for A.*

The proof is given in the Appendix. Working with Gaussian variables considerably simplifies operations, as independence between Gaussian variables is equivalent to decorrelation, and the minimum mean-squared error (MMSE) estimator—which minimizes D_n—reduces to a linear function in the Gaussian setting. As seen below, this implies that both f and g are linear—which is not too surprising as linear functions are known to preserve normality.

Notice that since $g(Y^i)$ is a sufficient statistic of Y_i for A, it does not matter if the feedback comes from the endpoints $\widehat{A}^i$

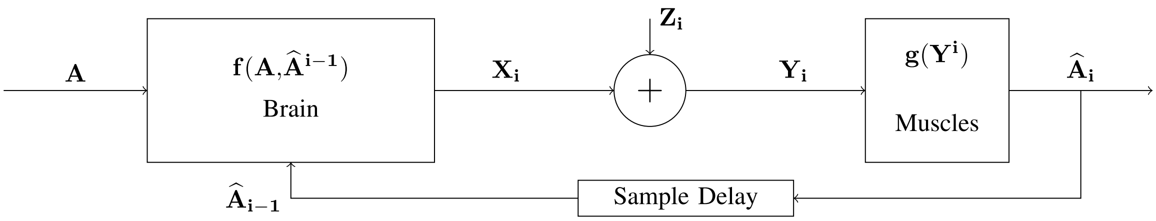

Fig. 4. Information-theoretic model for the aiming task, with initial distance A to the target at the end of the first phase; amplitude X_i created by the brain at $t = \tau + iT$; noisy version Y_i of X_i; distance $\widehat{A}^i$ actually covered by the limb at $t = \tau + iT$, where T is the infinitesimal time between two iterations.

or from the outputs of the channel Y^i. In this way our model can 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, namely the expression of f and g. We first obtain g by using the following result known as the *orthogonality principle*: if a certain quantity x is to be estimated from the observed data y by the unbiased estimator $\hat{x}(y)$, the orthogonality principle states that the following are equivalent:

- $\hat{x}(y) = \mathbb{E}[x|y] = \mathbb{E}[xy]^t \mathbb{E}[yy^t]^{-1} y$ is the unique MMSE estimator;
- $\mathbb{E}[(x - \hat{x}(y))y^t] = 0$.

Theorem 2. *For the optimal transmission scheme, $g(Y^i)$ is the MMSE estimator, $g(Y^i) = \mathbb{E}\big[A|Y^i\big]$.*

Proof. From condition 4 one has $\mathbb{E}[(A - g(Y^i))Y^i] = 0$. 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 $Y^i = (Y_1, ..., Y_i)$.

Theorem 3. *For the optimal transmission scheme,*

$$X_i = f(\widehat{A}^{i-1}, A) = \alpha_i (A - \widehat{A}_{i-1}) \tag{6}$$

$$= \alpha_i (A - \mathbb{E}\big[A|Y^{i-1}\big]), \tag{7}$$

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.

The previous two theorems formally define the encoding function f and decoding function g; both are linear functions. Incidentally, the motor system is known to be able to produce linear functions [36]. 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.

Theorem 4. *Let $A_i = X_i/\alpha_i$ be the unscaled version of X_i. We have*

$$\mathbb{E}[A|Y^{i-1}] = \sum_j^{i-1} \mathbb{E}[A|Y_j] = \sum_j^{i-1} \mathbb{E}[A_j|Y_j]. \tag{8}$$

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

Finally, we check optimality in Eq. (5) by evaluating the distortion in the following

Theorem 5. *The quadratic distortion $D_i = \mathbb{E}[(A - \widehat{A}_i)^2]$ decreases exponentially in i:*

$$D_i = \frac{\sigma_0^2}{(1 + P/N)^i}. \tag{9}$$

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 makes the correspondence between the transmission of one real value A using feedback, with that of n independent channel uses (transmissions).

It is important to note that the obtained equations were already given in an information-theoretic context by Gallager and Nakiboğlu [37] who discussed an older scheme by Elias [38]. To our knowledge, the constructive approach of the Elias scheme given here, as well as its application to model human aimed movements is novel.

V. Empirical Evidence

Let $t = nT$ be the total time spent during the second phase and notice that $D_n = \sigma_n^2$ is the variance of endpoints after n iterations. From Theorem 5, we have

$$\log_2 \sigma_n = \log_2 \sigma_0 - C/Tt = \log_2 \sigma_0 - C't. \tag{10}$$

where C is the Shannon capacity (3) and $C' = C/T$ is the capacity in bits/s. This equation is illustrated on real data in Fig. 5. The figure shows the standard deviation profile from participant 3 performing under the balanced (3) instruction. For $\tau > 0.18$ s, the logarithm $\log_2 \sigma(t)$ of the standard deviation decreases regularly and linearly with high goodness of fit ($r^2 = 0.984$, Student t-test for significance of the slope yields $F(1,76) = 4693, \quad p = 10^{-16}$). Computing the same statistics on the other 11 participants, we found an average goodness of fit of $r^2 = 0.93$ and an average $C' = 4.52$ bits/s. This provides strong empirical support for Theorem 5.

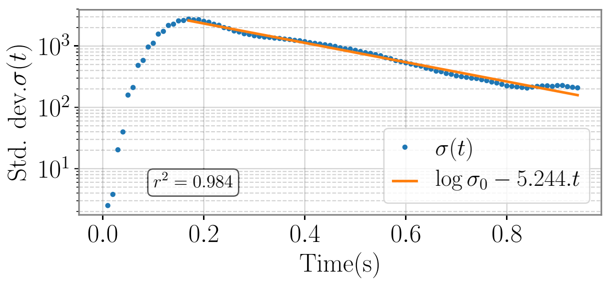

Fig. 5. Linear decrease of the logarithm of standard deviation of position over time for participant 3 of the G-dataset, performing in the balanced (3) condition.

VI. Conclusion

In this paper, we have established that for human aimed movements, the profiles of positional variance are unimodal in time, and that the variance-decreasing phase can naturally be modeled by an information-theoretic transmission scheme with feedback information from the current position. Positional variance decreases at best exponentially at a rate given by Shannon's capacity C. Increasing accuracy, i.e., decreasing the endpoint variance from σ_0^2 to σ_n^2, requires longer time t as shown in (10). This is consistent with the well-known speed-accuracy tradeoff[5]. We believe that there is a lot of insight to be gained from information-theoretic considerations in the study of human aimed movements. An interesting perspective is to consider the effect of non-ideal feedback or multiplicative noise on the optimal rate, and their practical implications.

Appendix A

Proof of Theorem 1. The proof uses well known techniques and inequalities from information-theory [24]. For inequality (4a):

$$\begin{aligned} I(A;\widehat{A}_n) &= H(A) - H(A|\widehat{A}_n) && (11)\\ &= H(A) - H(A-\widehat{A}_n|\widehat{A}_n) && (12)\\ &\geq H(A) - H(A-\widehat{A}_n) && (13)\\ &\geq H(A) - \frac{1}{2}\log\Big(2\pi e\mathbb{E}[(A-\widehat{A}_n)^2]\Big) && (14)\\ &= \frac{1}{2}\log\frac{\sigma_0^2}{D_n} && (15)\end{aligned}$$

Eq. (11) by definition of mutual information; Eq. (12) because of the conditioning by $\widehat{M}_n$; Eq. (13) because conditioning reduces entropy H; Eq. (14) because the Gaussian distribution maximizes entropy under power constraints; Eq. (15) by definition of the distortion and the entropy formula for a Gaussian distribution.

[5]In [34], we show that our result is also consistent with the well known Fitts' law.

For inequality (4b):

$$\begin{aligned} I(A;\widehat{A}_n) &\leq I(A;Y^n) && (16)\\ &= H(Y^n) - H(Y^n|A) && (17)\\ &= \sum_i \big[H(Y_i|Y^{i-1}) - H(Y_i|Y^{i-1},M)\big] && (18)\\ &= \sum_i \big[H(Y_i|Y^{i-1}) - H(Y_i|X_i)\big] && (19)\\ &\leq \sum_i [H(Y_i) - H(Z_i)] && (20)\\ &\leq \sum_i \left[\frac{1}{2}\log(2\pi e(P_i+N)) - \frac{1}{2}\log(2\pi eN)\right] && (21)\\ &\leq \sum_i \left[\frac{1}{2}\log\left(1+\frac{P_i}{N}\right)\right] \leq nC && (22)\end{aligned}$$

Eq. (16) by the data processing inequality [24], where $M \longrightarrow Y^i \longrightarrow g(Y^i) = \widehat{A}^i$ form a Markov chain; Eq. (17) by definition; Eq. (18) by applying the chain rule to both terms; Eq. (19) because of the feedback scheme; Eq. (20) because conditioning reduces entropy, $Y_i = Z_i + X_i$ and X_i and Z_i are independent; Eq. (21) because the Gaussian distribution maximizes entropy and X_i and Z_i are independent (where P_i and N are the powers of respectively X_i and Z_i); Eq. (22) by the concavity of the logarithm function.

□

Proof of Lemma 1. The proof consists of finding the conditions that turn the inequalities in the proof of Theorem 1 into equalities. Equality in Eqs. (13) and (16) directly imply condition 4. Equality in Eq. (14) implies that the $A - \widehat{A}^i$ are Gaussian. Equality in Eq. (16) 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 Eq. (20) implies condition 3. Equality in Eq. (21) implies that the Y_i's are Gaussian. Eq. (22) 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 3. 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 reads that $\mathbb{E}[\widehat{A}_i\widehat{A}^{i-1}] = \mathbb{E}[g(X_i + Z_i)\widehat{A}^{i-1}] = 0$. Because Z_i is independent from X_i and $\widehat{A}^{i-1}$, and g is linear, we get $\mathbb{E}[g(X_i)\widehat{A}^{i-1}] = 0$. Combining the two results, and because g is linear, we have $\mathbb{E}[\alpha_i(A - \widehat{f}(\widehat{A}^{i-1}))\widehat{A}^{i-1}] = 0$. The orthogonality principle clearly appears and $\widehat{f} = \mathbb{E}\big[A|A^{i-1}\big]$ is the MMSE estimator. □

Proof of Theorem 4. The goal is to evaluate $\mathbb{E}[A|Y^{i-1}]$. We first use the operational formula for the conditional expectation $\mathbb{E}[A|Y^{i-1}] = \mathbb{E}(AY^{i-1})^t\mathbb{E}[Y^{i-1}(Y^{i-1})^t]^{-1}Y^{i-1}$. We get, using the fact that the Y_j are independent and have power $P+N$:

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

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 5. 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. $\square$

Acknowledgment

We thank Yves Guiard who kindly provided access to his dataset used throughout this paper.

References

[1] R. S. Woodworth, "Accuracy of voluntary movement." *The Psychological Review: Monograph Supplements*, vol. 3, no. 3, p. i, 1899.

[2] E. Crossman, "The speed and accuracy of simple hand movements," *The nature and acquisition of industrial skills*, 1957.

[3] R. W. Soukoreff and I. S. MacKenzie, "Towards a standard for pointing device evaluation, perspectives on 27 years of fitts' law research in hci," *International journal of human-computer studies*, vol. 61, no. 6, pp. 751–789, 2004.

[4] P. M. Fitts, "The information capacity of the human motor system in controlling the amplitude of movement." *Journal of experimental psychology*, vol. 47, no. 6, p. 381, 1954.

[5] E. Todorov and M. I. Jordan, "Optimal feedback control as a theory of motor coordination," *Nature neuroscience*, vol. 5, no. 11, pp. 1226–1235, 2002.

[6] S.-C. Lai, G. Mayer-Kress, J. J. Sosnoff, and K. M. Newell, "Information entropy analysis of discrete aiming movements," *Acta Psychologica*, vol. 119, no. 3, pp. 283–304, 2005.

[7] M. A. Khan, I. M. Franks, D. Elliott, G. P. Lawrence, R. Chua, P.-M. Bernier, S. Hansen, and D. J. Weeks, "Inferring online and offline processing of visual feedback in target-directed movements from kinematic data," *Neuroscience & Biobehavioral Reviews*, vol. 30, no. 8, pp. 1106–1121, 2006.

[8] J. Van der Meulen, R. Gooskens, J. Denier Van der Gon, C. Gielen, and K. Wilhelm, "Mechanisms underlying accuracy in fast goal-directed arm movements in man," *Journal of Motor Behavior*, vol. 22, no. 1, pp. 67–84, 1990.

[9] W. Darling and J. Cooke, "Changes in the variability of movement trajectories with practice," *Journal of Motor Behavior*, vol. 19, no. 3, pp. 291–309, 1987.

[10] S. R. Gutman and G. L. Gottlieb, "Basic functions of variability of simple pre-planned movements," *Biological Cybernetics*, vol. 68, no. 1, pp. 63–73, 1992.

[11] R. A. Schmidt, H. Zelaznik, B. Hawkins, J. S. Frank, and J. T. Quinn Jr, "Motor-output variability: A theory for the accuracy of rapid motor acts." *Psychological review*, vol. 86, no. 5, p. 415, 1979.

[12] J. Soechting and F. Lacquaniti, "Invariant characteristics of a pointing movement in man," *Journal of Neuroscience*, vol. 1, no. 7, pp. 710–720, 1981.

[13] J. Soechting, "Effect of target size on spatial and temporal characteristics of a pointing movement in man," *Experimental Brain Research*, vol. 54, no. 1, pp. 121–132, 1984.

[14] W. A. Wickelgren, "Speed-accuracy tradeoff and information processing dynamics," *Acta psychologica*, vol. 41, no. 1, pp. 67–85, 1977.

[15] N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*. MIT press, 1948, 2nd revised ed. 1961.

[16] D. Elliott, R. Chua, B. J. Pollock, and J. Lyons, "Optimizing the use of vision in manual aiming: The role of practice," *The Quarterly Journal of Experimental Psychology Section A*, vol. 48, no. 1, pp. 72–83, 1995.

[17] H. N. Zelaznik, B. Hawkins, and L. Kisselburgh, "Rapid visual feedback processing in single-aiming movements," *Journal of motor behavior*, vol. 15, no. 3, pp. 217–236, 1983.

[18] R. Chua and D. Elliott, "Visual regulation of manual aiming," *Human Movement Science*, vol. 12, no. 4, pp. 365 – 401, 1993. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016794579390026L

[19] D. Elliott, W. F. Helsen, and R. Chua, "A century later: Woodworth's (1899) two-component model of goal-directed aiming." *Psychological bulletin*, vol. 127, no. 3, p. 342, 2001.

[20] D. Elliott, S. Hansen, L. E. Grierson, J. Lyons, S. J. Bennett, and S. J. Hayes, "Goal-directed aiming: two components but multiple processes." *Psychological bulletin*, vol. 136, no. 6, p. 1023, 2010.

[21] L. G. Carlton, "Visual processing time and the control of movement," *Advances in psychology*, vol. 85, pp. 3–31, 1992.

[22] J. Müller, A. Oulasvirta, and R. Murray-Smith, "Control theoretic models of pointing," *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 24, no. 4, p. 27, 2017.

[23] C. E. Shannon, "A mathematical theory of communication, part i, part ii," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, 1948.

[24] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.

[25] I. S. MacKenzie, "A note on the information-theoretic basis for fitts' law," *Journal of motor behavior*, vol. 21, no. 3, pp. 323–330, 1989.

[26] R. W. Soukoreff and I. S. MacKenzie, "An informatic rationale for the speed-accuracy trade-off," in *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on.* : IEEE, Oct 2009, pp. 2890–2896.

[27] J. Gori, O. Rioul, and Y. Guiard, "Speed-accuracy tradeoff: A formal information theoretic transmission scheme (FITTS)," *ACM Transactions on Human Computer Interaction*, To appear, available online at https://hal.archives-ouvertes.fr/hal-01690089/document.

[28] R. D. Luce, "Whatever happened to information theory in psychology?" *Review of general psychology*, vol. 7, no. 2, p. 183, 2003.

[29] Y. Guiard, H. B. Olafsdottir, and S. T. Perrault, "Fitt's law as an explicit time/error trade-off," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2011, pp. 1619–1628.

[30] P. M. Fitts and J. R. Peterson, "Information capacity of discrete motor responses." *Journal of experimental psychology*, vol. 67, no. 2, p. 103, 1964.

[31] J. A. Hartigan and P. M. Hartigan, "The dip test of unimodality," *The Annals of Statistics*, pp. 70–84, 1985.

[32] D. E. Meyer, R. A. Abrams, S. Kornblum, C. E. Wright, and J. Keith Smith, "Optimality in human motor performance: Ideal control of rapid aimed movements." *Psychological review*, vol. 95, no. 3, p. 340, 1988.

[33] P. A. Hancock and K. M. Newell, "The movement speed-accuracy relationship in space-time," *Motor Behaviour: Programming, Control, and cquisition*, pp. 153–188, 1985.

[34] J. Gori and O. Rioul, "A feedback information-theoretic transmission scheme for modelling aimed movements," In preparation.

[35] G. P. van Galen and W. P. de Jong, "Fitts' law as the outcome of a dynamic noise filtering model of motor control," *Human movement science*, vol. 14, no. 4, pp. 539–571, 1995.

[36] D. A. Rosenbaum, *Human motor control*. Academic press, 2009.

[37] R. G. Gallager and B. Nakiboğlu, "Variations on a theme by schalkwijk and kailath," *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 6–17, 2010.

[38] P. Elias, "Channel capacity without coding," in *Proceedings of the Institude of Radio Engineers*, vol. 45, no. 3, 1957, pp. 381–381.