Abstract

Many dynamic and kinematic features of reaching movements present invariants depending on the context of the task at hand. Two basic phenomena interact in the way the speed of ~~our ~~reaching movements is determined. First, we tend to reach faster a target that looks more rewarding, despite the additional muscular cost of a faster movement. Second, when we need to be more ~~precise~~accurate, ~~our ~~movements take~~s~~ more time. So far, these two phenomena have been studied in isolation despite their obvious interdependency. ~~In particular, two r~~Recent computational approaches ~~models of motor control address the first phenomenon. They ~~suggested~~explain~~ that the emergence of ~~the time of ~~movement duration ~~as ~~result~~ing~~ed from a cost-benefit trade-off arising from the summation of a temporally discounted reward and a cost that increases for faster movements. However, these models ~~do ~~did not account for the ~~second phenomenon, i.e. the ~~dependency between movement time and precision requirements, resulting in a speed-accuracy trade-off and formally expressed through~~by~~ Fitts' law. Another model ~~addresses ~~addressed the role of this speed-accuracy trade-off in determining movement time, but ~~does ~~did not take the cost of movement into account. ~~In this paper~~Here, we propose a framework that unifies the cost-benefit ~~trade-o ~~and~~ the~~ speed-accuracy trade-offs to explain movement properties related to time. With respect to the cost-benefit trade-off models, precision constraints are incorporated through the derivation of a new optimization criterion that considers probabilistic reaching of a rewarding target that may be missed if the motion is too fast. Using this computational model, we investigate the more global trade-off arising from the interactions between movement time, cost and accuracy. We show that this model accounts for Fitts' law and for other well-established results in the motor control literature.

1. Introduction

When grasping a hot cup of coffee on a table, our brain must take decisions upon values of myriad parameters that characterize the movement and the way our body interacts with its surrounding environment. On the one hand, it might be detrimental to go too fast because a collision between the hand and the cup could occur and hence, harm the skin. On the other hand, if it takes too long… the coffee will be cold. A specific control policy is selected by compromising the values of parameters in a cost function such as reward, muscular effort and movement duration.

There has been a recent progress in motor control research on understanding how the time of a reaching movement is chosen. In particular, two recent models from Shadmehr et al. [1] and Rigoux & Guigon [2] proposed an optimization criterion that involves a trade-off between ~~the~~ muscular effort and ~~the~~ subjective value of getting the reward, hence a cost-benefit trade-off ~~(CBT)~~. On the one hand, reaching a target faster requires a larger muscular effort (~~refs?~~Berniker et al., 2013). On the other hand, the subjective ~~value~~ utility of reaching a target decreases as the time needed to reach the target is increased (Haith et al., 2012~~refs?~~). Consequently, ~~As a result,~~ the net expected return resulting from ~~consisting of~~ the difference between subjective ~~value~~ utility ~~minus the~~and muscular effort is optimal for a certain time, as illustrated in Figure~~.~~ 1(A).

However, these models ~~do~~ did not account directly for basic facts about the relationship between movement difficulty and movement duration as captured more than fifty years ago by Fitts' law [3]. According to this law, the smaller a target, the slower the reaching movement. This is well ~~explained~~ captured by the so-called speed-accuracy trade-off ~~(SAT)~~ stating that, the faster a movement, the less accurate it is, hence the higher the probability to miss the target. Therefore, when accuracy is a constraint, ~~So a subject reaching too fast may not get the subjective value associated to~~ reaching movements ~~and~~ should slow down.

In contrast with the models of [1] and [2], the ~~model of~~ Dean’s model [4] takes the speed-accuracy trade-off~~SAT~~ into account with ~~. T~~the key difference ~~with respect to [1] and [2] is~~ that, instead of maximizing a reward, ~~this model~~it maximizes a reward ***expectation***. In other words, it maximizes the ~~, i.e. the~~ reward multiplied by ~~times~~ the probability to get it.

However, the model proposed in [4] is an abstract model of movement time selection that looks for an optimal trade-off between an externally decayed reward and a speed-accuracy trade-off ~~SAT~~ that relates the probability of missing to movement time. As such, it does not account for movement execution, neither for the ~~choice~~ planning of a motor trajectory and its impact on the cost of movement. Furthermore, ~~The~~ the model does not explain Fitts' law~~,~~; it rather incorporates its consequences into an abstract model of the ~~SAT~~speed-accuracy trade-off that is fitted to experimental data. The mathematical design of the model is based on several simplifying assumptions and it predicts optimal movement times that are systematically shorter than those observed with subjects. The authors of [4] discuss that this may result from the fact that the model does not take the cost of movement into account.

In this paper, we show that the models of Shadmehr [1] and Rigoux [2] as well as the model of Dean [4] can be unified into a model that solves the difficulties faced by these previous models.

This unification is simply implemented by including sensory and motor noise into the optimal control model proposed in [2], shifting from a deterministic account of the movement to a stochastic one, in line with other modeling approaches ~~the models of~~ [5-9].

As a matter of fact, in the models of [1] and [2], the target is given as a single point and the movement is considered as always reaching it, irrespective of the size of the target. In order to fully account for Fitts' law, one must consider the intrinsic dispersion of reaching movements towards a target and the effect of sensory and muscular noise on this dispersion (e.g. [5], see [10] for a review), which is not the case of the models of [1] and [2].

As a result, the reward and muscular activation terms in the optimization criterion proposed in [2] are replaced by reward and cost expectation terms. Considering expectation is a way to account for the fact that, in case of a miss, one would not get the reward, so the global outcome of the movement would only consist of its incurred cost.

The mathematical way to capture this intuition is presented in Section 4.1 and illustrated in Fig. 1~~(~~B~~)~~. With respect to Fig. 1~~(~~A~~)~~, the new model includes an additional term that expresses the dependency between the size of the target and the probability to hit it, as expressed by the speed-accuracy trade-off~~SAT~~. As can be seen in Fig. 1~~(~~B~~)~~, if the target is smaller, then the probability to hit ~~get~~ it is smaller for a given time, thus the reward expectation should itself be smaller. As a result, the optimum time resulting from the optimal combination of this reward expectation with the cost of movement should shift to longer times, which is qualitatively consistent with Fitts' law.

Beyond a theoretical motor control model, this paper proposes a computational model that is tested against a set of experimental phenomena.

2 Results

The goal of the computational studies hereafter was to investigate the behavioral properties of the proposed model in order to highlight the differences with respect to [1], [2] and [4].

In a first step we illustrate some basic motor control properties of the model at the level of trajectories and motor cost. In a second step, we examine the complex relationship between movement time, final dispersion and the expected gain arising from this model. In a third step, we show that the model globally accounts for Fitts' law. Finally, we show that it generates an asymmetric velocity profile, where the asymmetry increases with the amplitude of signal-dependent motor noise.

3.1 Movement cost is not a symmetric function of target location

~~In order to investigate basic properties of the motor control model, it is demonstrated on a~~We simulated two-joint planar arm reaching movements with six muscles with the same model as [REF]. ~~taken from the literature and illustrated in Fig. 9.~~ This simulated arm model is shown in Fig. 9 and described in more details in Section 4.3.

Figure 2 depicts ~~The~~ the cost obtained from the different initial configurations. ~~described in Section 4.3.4 is shown in Figure 2.~~

As expected, one can observe that the smaller the distance to the target, the lower the cost (red and yellow color). Furthermore, starting from the ~~left~~ left-hand~~-~~ side of the goal point results in a lower cost than starting from the ~~right~~ right-hand~~-~~ side. This is explained by the fact that the optimal muscular strategy for performing these movements differs depending on the side. Actually, movements starting from the right are performed by moving simultaneously the elbow and the shoulder whereas when starting from the left, only the shoulder is involved, leading to a lower motor cost. The muscular activations corresponding to these two situations are shown in Figure 3.

One can see that the pattern of activation varies depending on the initial configuration of the movement, corresponding to two different optimal strategies: when starting from a short distance to the ~~left~~ left-hand~~-~~ side, only the shoulder is moved, whereas when starting from a long distance to the ~~right~~ right-hand~~-~~ side, both the elbow and shoulder are moved, resulting in a more complex muscular activation strategy. One can also see that co-contractions ~~is~~ are avoided, consistently with the minimum intervention principle [6].

3.2 Relations between movement time, expected gain and final dispersion

The model presented ~~in this paper~~here is designed to investigate the complex relationships between movement duration, expected gain and final dispersion. In order to perform this investigation, we chosed a set of five target sizes (~~f~~1mm, 2mm, 6mm, 10mm, 20mm~~g~~). For each (distance, target size) pair, we optimized a specific controller (see Methods). For each of these controllers corresponding to each (distance, target size) pair, we recorded movement time, final dispersion and performance. The corresponding data is shown in Fig. 4.

As illustrated in Fig. 1~~(~~B~~)~~, the probability to reach the target depends on the size of the target and the time of the movement (or its velocity). More precisely, if the target is smaller, fast movements should fail more often. Thus, as a result of including the accuracy constraint in the model, the optimal movement time resulting from the intuition described in Fig. 1~~(~~B~~)~~ should be always longer than the optimal movement time resulting from the model of [2].

This is what is observed in Fig. 4~~(~~A~~)~~. One can see that movement time increases when the target is smaller, and also increases with the movement distance, consistently with Fitt's law.

~~In Fig. 4(B), o~~We also observe that ~~ne can see that~~ the net expected return is smaller for a larger distance (Fig. 4B), because the muscular effort for performing a larger movement is larger.

Most importantly, the net expected return increases with the size of the target. There are two explanations for this fact. First, it means that the benefit in terms of subjective value from reaching the target faster is higher than the increase in cost resulting from a faster movement. Second, if the target is larger, less precision constraints on the movement can result in a better trajectory in terms of muscular activations to reach the target.

Finally, Figure 5 shows an example of the obtained dispersion corresponding to the five targets from a distance of 18cm, using 100 trajectories from the same starting point at x = 0 to the target.

One can observe that, when the target is smaller, dispersion is reduced to increase the probability of reaching the target. In order to reduce dispersion, the motion is slowed down~~performed slower~~, as illustrated in Fig. 4A. However, even with very slow movements, ~~for very small targets,~~ the probability to miss small target ~~to miss the target~~ is not zero~~null~~. This means that, with our parameter settings, it is more optimal to pay the price of a few failed movements than to move slow enough to succeed at all times. There are even target sizes for which reaching may fail whatever the movement velocity [11].

3.3 Reproduction of Fitts' law

Fitts' law states that movement time (MT) is linear in its difficulty index (DI), this index being bigger for longer movements and smaller targets. Fitts' law is written:

$$MT = a + b.\log_2 \underbrace{\left(\frac{D}{W}\right)}_{DI}$$

where D is the distance of the movement~~ (denoted with A for amplitude in other papers)~~, W is the width of the target and a and b are linear coefficients. This law was initially studied for one dimensional movements, and then extended ~~for~~ to many other contexts [12-17].

~~From the data presented in Section 2.2 and using (1), w~~We computed DI values for different distances D and target widths W from the data presented in Section 2.2 and using (1)~~-~~. Figure 6 shows the ~~resulting~~ movement times (MT) against ~~over~~ Difficulty Indexes (DI).

~~One can see that we get a~~A clear linear relationship exists between these two variables~~s~~, which demonstrates that ~~thus the data~~our model is consistent with Fitt's law.

The obtained values of a and b cannot be compared to empirical data from the human motor control literature given the wide variability of these values ~~accross~~across subjects [18,19].

3.4 **Velocity profiles**

The final dispersion in reaching trajectories is generated by motor noise. Following the minimum intervention principle from [6], motor noise being proportional to muscular activation (signal-dependent), the only way to decrease motor noise is to decrease muscular activation.

Thus, in order to hit a small ~~trajectory~~target, muscular activations should be small when the reach approaches the ~~by the~~ end of the movement, which can result in first instance in less co-contraction and then in smaller ~~less~~ velocity. Furthermore, a slower movement provides a better opportunity for state estimation to compensate for delayed feedback about the position of the end effector. Taken together, those two phenomena contribute to the fact that an optimal controller should generate less velocity by the end of the movement for a smaller target. Put in another way, the probability to hit a target is maximal if ~~So one way to make sure to hit a small target would be to~~we perform a slow reaching movement.

However, as explained above, a slower movement results in a discounted reward, thus the movement should nevertheless be as fast as possible.

As a consequence, the best option for optimizing reaching accuracy under temporal constraints consists in being very fast in the beginning of the movement and much slower in the end. Thus the velocity profile should be asymmetric. The main drive for this asymmetry being motor noise, the more motor noise, the more asymmetric the movement should be.

This is what we observe in Figure 7, where velocity profiles are generated for different amplitudes of motor noise.

Incidentally, one can observe on the ascending parts of the profiles that we do not get strictly the shape of a bell curve. This is due to the limited optimization capability of our methods, given the constraint on the number of samples.

3 Discussion

3.1 ~~Positionning~~Positioning

When performing a reaching movement towards a target, three inter-related factors must be determined. The first is the potential outcome of a successful or failed movement, characterized by a discounted reward. Movement ~~time~~ duration is crucial in this factor because more time means more discount. The second factor is the cost of the movement, depending on muscular activations and resulting in a velocity profile and a joint-space trajectory. Movement ~~time~~ duration is again central ~~is crucial~~ in this factor because, for an identical trajectory, less time means more cost. The third factor is the final dispersion~~,~~ generated by motor noise and that is imperfectly compensated for by the Central Nervous System~~CNS~~ due to sensory noise and feedback delays. Movement duration ~~time~~ is also ~~crucial~~ critical to tune this ~~in~~ this factor because less time means more dispersion.

The works of [2] and [1] only relate the first two of these factors. By contrast, the model presented in [4] relates the first and the last factors, without consideration for the second.

The model presented here ~~in this paper~~ addresses the more global inter-relationship between these three factors and provides an optimality criterion that accounts for the strategy of human subjects in this multi-dimensional choice space.

3.2 Background

Our ~~The~~ model ~~presented in this paper~~ is consistent with the stochastic optimal control view of motor control [5~~{~~-9]. It starts from the fact that, for a small target, the faster the movement, the lesser the chance to hit it (refs). According to this view, there are two complementary explanations for this fact. First, the motor activation signal descending from the Central Nervous System ~~(CNS)~~ to motoneurons is corrupted with some noise that is proportional to this signal (refs). Thus a faster movement means more noise, hence more intrinsic dispersion of the final hit point if the arm was controlled in a purely open-loop way.

Second, this intrinsic tendency to dispersion is compensated for by a feedback control loop which is based on state estimation mechanisms. State estimation itself is based on delayed proprioceptive and perceptive feedbacks. Thus a faster movement means less time to accurately estimate the current state, hence less compensation for dispersion. As a consequence of both mechanisms, a faster movement results on more final dispersion, hence in a lesser probability to hit a small target.

3.3 **Contributions**

The model proposed here takes the one of [2] as a starting point. This former model reproduces basic characteristics of motor behavior, as expected from the close relationship with previous optimal control models [6, 20-~~{~~24]. It also explains several phenomena in cost-benefit trade-off tasks [25, 26]. The model presented here is equivalent to the one presented in [2] when the probability to hit the target is set to 1. As a result, it still benefits from the above-~~–~~mentioned properties that are not impacted by this probability. In this section we show that it solves limitations of the models of [4] and [2].

3.3.1 *The model accounts for target selection bias*

In [27], the authors mention a systematic pointing bias in both the x and y directions for all subjects performing a target hitting experiment. The presence of this bias contrasts with an assumption made in the model of [4] that, spatial errors being symmetric, the optimal choice of x and y should be in the middle of the circular target.

The proposed model shows that this simplifying assumption does not hold. Indeed, as illustrated in Figures 2 and 3, the optimal strategy to reach a target significantly changes depending on the relative position of the starting point and the target point. As a consequence, when aiming at a large target in the center of the ~~saggital~~sagittal plane in front of himself/herself, an optimal subject would not aim at the center of the target. The optimal aiming point depends on the function relating movement cost to aiming point location, which itself depends on the musculoskeletal system of the subject. This explains why the bias differs from one subject to the other.

Figures 2 and 3 are obtained in a 2D case with a 1D target whereas Dean's experiments were performed in 3D for a circular target. Nevertheless, the asymmetry resulting from the proposed model would also be present in a 3D model.

3.3.2 *The model directly accounts for Fitts' law*

Motor control results are obtained in [2] in the absence of sensory and motor noise. As such, this model cannot provide a direct account of phenomena relying on the stochasticity of the motor system, such as Fitts' law. Actually, the model of [2] provides an indirect account of Fitts' law (see [2], Fig. 7A). For obtaining these results, the authors have estimated dispersion as a function of velocity considering a constant velocity over the movement, and they have reconstructed the relationship between Difficulty Index and movement time based on the size of a target that would match this estimated dispersion ([2], personal communication). So Fig. 7A in [2] is based on one target size only.

In [4], an abstract ~~SAT~~ speed-accuracy tradeoff model is directly fitted to human movement data, without directly calling upon~~d~~ a measured movement dispersion.

In contrast, in Section 2.3 we have shown that the proposed model accounts for Fitts' law by using several targets and several starting points. In this model, movement velocity is far from constant and dispersion is measured as an effect of motor noise and imperfect state estimation rather than inferred based on an a priori speed-accuracy tradeoff~~SAT~~ model.

3.4 **Limitations**

3.4.1 *The model does not account for movement planning*

In [27,28], the authors distinguish movement planning from motor planning. Movement planning consists in choosing where reaching should aim given a set of rewarded and penalized targets and motor variability. By contrast, motor planning consists in specifying movement execution in advance, in terms of muscular activations at each step of the movement, given a chosen target. Movement planning does not take motor costs into account and does not account for movement time. Interestingly, the work of [27,28] is focused on movement planning, thus it does not account for the motor trajectory and the choice of movement time as the models of [1] and [2] do.

The model proposed here might be seen as providing a first stone of the bridge between the work of [27,28] and the one of [1] and [2].

The model presented in this paper cannot explain the capability to immediately combine information about these interacting targets when they are visible, as reported in [29]. Some inference mechanism must be assumed to explain this immediate composition capability. At least part of this inference is probably initiated before movement execution starts. More generally, there is no mechanism in the model proposed here to account for movement preparation (e.g. [30]), though this stage certainly plays a role in the phenomena studied here.

3.4.2 *Muscular effort or activations?*

It has been shown (ref) that using jju2jj or muscular effort or... results in very similar movements.

3.4.3 *Exponential versus hyperbolic discounting*

The proposed model starting from the one in [2], it inherits from this model an exponential discounting of the reward through time. In an alternative model, [1] rather suggests an hyperbolic discounting approach, in line with many other authors (e.g. [31]). At this stage, we consider that the debate between diverse discounting approaches is far from closed (see e.g. [32]) and using a different discounting approach would not fundamentally change the results presented in this paper.

3.4.4 *Expectation over reward or expected gain*

The proposed model computes the expectation over the reward part rather than on the sum of the reward and the movement cost. The intuition behind this choice is that the reward term varies a lot depending on whether the target was hit or not whereas the movement cost is grossly constant over movements from the same point to a same target. If the movement cost was actually constant over movements, it could be left out of the expectation term without harm. To discriminate between both potential models, one should investigate experimental settings where the cost of movements varies a lot, for instance using force fields. This is left for future work.

3.4.5 *Imperfect optimization*

Results in Figures 4(C) and 7 show that the incremental optimization process used in this paper (see Section 4) was not given enough iterations to reach a global optimum.

3.5 **Predictions**

The computational study presented in this paper can be seen as generating a number of predictions that remain to be tested experimentally.

First, in the context of planar reaching movements towards a large target in front of the subject, we predict a tendency to move the distribution of hit points to the left, because movements towards the left are less expensive than movements along the ~~saggital~~sagittal plane.

Second, the model proposed here progressively optimizes its distribution of hit points based on the gain resulting from previous hits. Actually, the model proposed here addresses a situation that is quite different from the one experimented in the work of Trommersh•auser et al. [27, 28, 33]. Their work con-siders a situation where the subject can see the target and decides where to aim based on this available information. Decision is described as an inference process based on global information. In our model, by contrast, the search for the right hit point dispersion is a local trial-and-error process. Pre-training orients the controller towards an initial distribution of hit points, then the optimization process adapts this controller to a specific target but the controller is not given any prior information about the size or location of this target. It is only informed whether the target was hit or not through the reward feedback.

Thus, experimentally, the model presented in this paper would correspond to a situation where a subject is vaguely informed about the location of the target but has to adapt its reaching movement to ~~maximise~~maximize the outcome through trial-and-error. To our knowledge, this situation has never been studied experimentally in the ~~litterature~~literature.

The most closely related situation is the one described in [34], where the target is progressively shown to the subject by plotting more and more random points drawn according to the spatial distribution of the reward. Thus, in a way, the subject discovers the target through time, rather than through trial-and-error.

In the situation corresponding to the proposed model, it would be interesting to determine experimentally the circumstances under which a subject sacrifices accuracy depending on the target location, its size, its rewarding value and timing constraints over the movement. All the corresponding data could be checked against the predictions of the proposed model. In particular, one can anticipate that, if the reward gets null after a short time, subjects should perform the movement very fast at the expense of accuracy, given that hitting a rewarded target only part of the time is still better that receiving no reward at all over all trials.

3.6 From motor control to motor learning

The perspective taken here about our model consisted in considering the proposed methods as a tool to get optimal ~~behaviours~~behaviors with respect to the cost function defined by Eq. (3).

By the way, this method optimizes a parametric controller for a given target size and location by trial-and-error, without knowing these size and location in advance. For a particular context, it empirically optimizes the trade-off between cost and accuracy by tuning the motor input so that velocity generates the optimal dispersion for the given target. In that respect, the model might be considered under a motor learning perspective that would try to explain how we may learn optimal reaching movements from trial-and-error, but this is beyond the scope of this paper.

4 Material and methods

In the first part of this section, we describe the theoretical background of the model. In a second part, we derive ~~describe how we obtain~~ a computational model that optimizes the cost function described in ~~(~~[3~~)~~] for different contexts. Finally, the simulated arm and experimental apparatus used to model reaching are described in Section 4.3.

4.1 Mathematical formulation of the model

The cost function J(u) proposed for a control u in the model of [2] is

$$J(\boldsymbol{u}) = \int_0^\infty e^{-t/\gamma}[\rho R(\boldsymbol{s}_t) - \nu L(\boldsymbol{u}_t)]dt$$

where R(st) is the immediate reward function that equals 1 at the goal point (also called rewarded state) and is null everywhere else. The function L(ut) is the movement cost. In [2], t~~T~~he authors ~~of [2]~~ take L(ut) = kutk2, as in many motor control models. The continuous-time discount factor accounts for the “~~\~~greediness" of the controller, i.e. the smaller~~,~~ gamma, the more the agent is focused on short term rewards. Finally, rho is the weight of the reward term and eta the weight of the effort term. In all experiments presented here, based~~,~~ on the previous work from [2], we took gamma = 0:998, rho = 1 and eta = 3000.

A near optimal deterministic policy to solve this problem is obtained through a computationally expensive variation calculus method (see [35] for details). Given that the policy does not take the presence of noise in the model of the plant into account, the actions must be computed again at each time step depending on the new state reached by the plant which further contributes to the cost of the method. The controller resulting from this model is called the NOPS (for Near-Optimal Planning System) in the rest of this paper.

Now let us consider the integration of accuracy constraints. Instead of a deterministic controller, the new model is based on a stochastic controller where the rewarded state is reached or not. As a result, the outcome of a large set of movements performed with noise is computed as the value of the reward ~~mutiplied~~multiplied by the probability to obtain it over the different movements. Mathematically, the value multiplied by the probability is called the expectation.

Taking the probability to reach the target into account as described above, the new optimization criterion is written

$$J(u) = \int^{\infty} e^{-t/\gamma}\mathbb{E}[\rho R(s_t) - \nu L(u_t)]dt$$

where IE[] stands for the expectation of the cumulated reward, and R(st) equals 1 if the end effector hits the target.

4.2**Incremental stochastic optimization**

The optimal control problem arising from a cost function including a reward expectation cannot be solved analytically. The reward expectation itself must be estimated empirically through a set of attempts to hit the target (these attempts are called “roll-outs" hereafter). The more roll-outs, the better the estimate of the reward expectation. In [2], the simpler optimal control problem was solved with a numerical variation calculus method called the ~~\Near-Optimal Planning System" (nops) hereafter~~NOPS. This method is ~~computationnally~~computationally expensive as ~~,~~ it takes about 10 minutes for generating one reaching trajectory on a standard computer. As a consequence, it cannot be used as such to empirically determine the reward expectation for a given problem configuration.

To circumvent this difficulty, the computational model presented in this paper relies on a two-step approach. First, we approximated the ~~nops~~ NOPS using a nonlinear function approximation technique named XCSF which consists in ~~xcsf.xcsf is~~ a regression algorithm that can approximate a function in a large continuous space [36,37]. It generates a parametric model of the approximated function as a Gaussian mixture of linear models, i.e. a collection of local linear models bound to Gaussian support functions. A more complete description of ~~xcsf~~ XCSF can be found in [38~~{~~-40]. The result is a parametric controller that approximates the function relating the state s of the system to the adequate control input u the ~~nops~~ NOPS would provide in that state. This approximated function is called the XCSF ~~xcsf~~ controller. It is trained by using trajectories generated by the ~~nops~~ NOPS as input samples, using the cost function described in (2). Its parameters are the weights of all local linear models learned with XCSF.~~xcsf.~~

As described in [41], this controller mimics the NOPS ~~nops~~ in the limited region where it has been trained, but it reacts several orders of magnitude faster because the result of the ~~optimisation~~optimization process is ”compiled" into the controller parameters. From this much faster controller, it becomes possible to empirically estimate a discounted reward expectation from many roll-outs.

In a second step, the XCSF ~~xcsf~~ controller is re-optimized with respect to the cost function (3) for different target sizes and different sets of initial positions. This optimization is performed using a variant of the Cross-Entropy Method ~~(cem)~~ [42] illustrated in Fig. 8.

Given the initial XCSF ~~xcsf~~ controller learned from ~~nops~~ NOPS demonstrations, the method consists in optimizing the parameters of this controller by a local stochastic search method. New roll-outs are performed with varying parameters for all local linear models around those of the current controllers, and the parameters that give rise to a better performance with respect to the cost function (3) are retained in the new current controller. For more details about the methods, see [41].

We used the JavaXCSF [43] implementation of XCSF~~xcsf~~, and the main code for the experiments as well as the ~~ϵ~~CEPS~~eps~~ algorithm ~~are~~ were also implemented in Java. The simulations ~~experiments~~ were~~are~~ run on an Intel Core 2 Duo E8400 @ 3 GHz with 4 GB RAM.

XCSF ~~xcsf i~~was tuned as follows. The maximum number of local linear models (population) ~~is~~ was set to 200. Learning ~~is~~ was stopped after 200~~;~~ ,000 iterations. The input were~~are~~ normalized: the target and current positions ~~are~~ were bounded by the reachable space and the speed ~~is~~ was bounded by ~~[~~ [-100; +100] rad.~~s~~ s-1. The default action udefault ~~is~~ was set to a vector of zeros i.e., no muscular activation. After tuning empirically the parameters, the learning rate beta ~~is~~ was set to 0~~:~~.1, the accuracy factor alpha wa~~i~~s set to 1~~:0~~ and the deletion threshold delta ~~is~~ was set to 0~~:~~.1. Compaction, randomization and multithreading ~~are~~ were disabled to improve reproducibility of the results [43].

4.3**Arm model and experimental apparatus**

There are several models in the literature that combine a simple two joint planar rigid-body dynamics model with a muscular actuation model. Most of these models [9,44,45] are defined in the sagittal plane and ignore gravity effects, an exception being [46] that lies in the vertical plane and takes the ~~gravity~~ gravitational force into account.

Apart from this exception, the differences between the models above mostly lie in the muscular actuation component.

Our model is also a two joints planar arm in the ~~saggital~~sagittal plane controlled by 6 muscles, illustrated in Fig. 9, where the muscular actuation model is taken from [44] (pp. 356-357) as cited by [45].

Figure 9 shows the arm with the ~~positionning~~positioning of the muscles. Table 1 in Appendix A ~~reminds~~ reports the nomenclature of all the parameters and variables of the model.

4.3.1 Rigid-body dynamics

The rigid-body dynamics equation of a mechanical system is:

$$\ddot{q} = M(q)^{-1}(\tau - C(q, \dot{q})\dot{q} - g(q) - B\dot{q})$$

where q is the current ~~articular~~ joint position, q_ the current joint ~~articular~~ speed, q□ the current joint ~~articular~~ acceleration, M the inertia matrix, C the Coriolis force vector, the segments torque, g the gravity force vector and B a damping term that contains all unmodelled effects.
Here, g is ignored since the arm is working in the sagittal plane.
All angles are expressed in radians.

Given the arm parameters m1 = 1:4; m2 = 1:1; l1 = :30; l2 = :35, s1 = :11; s2 = :16; d1 = :025; d2 = :045, where mi is the mass of segment i, li the length of segment i, si the inertia of segment i and di the distance from the center of segment i to its center of mass, the inertia matrix can be computed as $M = \begin{bmatrix} k_1 + 2k_2\cos(q_2) & k_3 + k_2\cos(q_2) \\ k_3 + k_2\cos(q_2) & k_3 \end{bmatrix}$, with $k_1 = d_1 + d_2 + m_2 l_1^2,\ k_2 = m_2 l_1 s_2,\ k_3 = d_2$.

The Coriolis force vector is given by $C = \begin{bmatrix} -\dot{q}_2(2\dot{q}_1 + \dot{q}_2)k_2\sin(q_2) & \dot{q}_1^2 k_2 \sin(q_2) \end{bmatrix}$.

The damping term is $B = \begin{bmatrix} .05 & .025 \\ .025 & .05 \end{bmatrix}\dot{q}$.

The computation of the torque tau exerted on the system given an input muscular actuation u is explained in the next section.

4.3.2 *Muscular actuation*

Our muscular actuation model is taken from [44]. It is a simplified version of the one described in [9] in the sense that it uses a constant moment arm matrix A whereas [9] is computing this matrix as a function of the state of the arm.

From [44], we take the maximum force exerted by each muscle as

$$\begin{aligned} f_{\max} &= \begin{bmatrix} f_{\max 1} & f_{\max 2} & f_{\max 3} & f_{\max 4} & f_{\max 5} & f_{\max 6} \end{bmatrix} \\ &= \begin{bmatrix} 700 & 382 & 572 & 445 & 159 & 318 \end{bmatrix}^\top \end{aligned}$$

and the moment arm matrix is

$$\begin{aligned} A &= \begin{bmatrix} A_{11} & A_{21} & A_{31} & A_{41} & A_{51} & A_{61} \\ A_{12} & A_{22} & A_{32} & A_{42} & A_{52} & A_{62} \end{bmatrix}^\top \\ &= \begin{bmatrix} .04 & -.04 & 0 & 0 & .028 & -.035 \\ 0 & 0 & .025 & -.025 & .028 & -0.35 \end{bmatrix}^\top . \end{aligned}$$

Finally, given an action u corresponding to a raw muscular activation as output of the controller, the muscular activation is augmented with Gaussian noise using $\tilde{u} = \log(\exp(\kappa \times u_t \times (1 + \mathcal{N}(0, I\sigma_u^2))) + 1)/\kappa$, where $\times$ refers to the element-wise multiplication, I is a 6×6 identity matrix. and $\kappa = 25$ is the Heaviside filter parameter, and the input torque is computed as $\tau = A^\top(f_{\max} \times \tilde{u})$.

Given (4), the simulator uses the Euler method to compute the evolution of the system, with a time step of $\Delta t = 2$ ms.

4.3.3 *Motor noise and state estimation*

In order to reproduce Fitts' law and to study the structure of movement variability, taking sensory and motor noise into account is necessary. In order to get an adequate dispersion, we considered multiplicative motor noise wt with p(w) N (0; 0:4) and additive sensory noise t with p( ) N (0; 0:0004). The feedback delay is 100ms. State estimation was performed with an extended Kalman filter as described in [47].

4.3.4 *Experimental apparatus*

In most experiments where a subject has to reach a target, either in monkeys (e.g. [48]) or humans [4,27,29,34,49], the target is displayed on a vertical screen and the subject performs a trajectory in the sagittal plane that the screen intercepts. Our experimental apparatus reproduces such a scene in the sagittal plane.

Figure 10 shows the experimental apparatus for reaching. Cartesian coordinates (in meters) are expressed with respect to the position of the shoulder, taken as origin (x = 0; y = 0). The arm is shown together with the screen and the set of starting positions of the tested movements. In all ~~experiments~~ simulations, the target ~~is~~ was located at the same place, but its size ~~changes~~changed. The starting configuration ~~i~~was~~s~~ varied systematically over all movements so as to investigate the effect of the distance travelled ~~through~~ during the movement.

From a machine learning and control point of view, the state-space consists of the target ~~articular~~ joint position q , the current joint ~~articular~~ position q of the arm and its current joint ~~articular~~ speed q. The state s = (q ; q; q) has a total of 6 dimensions. The initial state is defined by a starting position, a null speed and the target position. The positions are bounded to represent the reachable space of a standard human arm, with q1 2 [0:6; 2:6] and q2 2 [0:2; 3:0], as shown in Fig. 10. The action-space consists of an activation signal for each muscle, which also makes a total of 6 dimensions.

To define a reaching movement with the ~~nops~~NOPS, a goal point is required. This goal point, shown as a red star in Fig. 10, is located at (x = 0; y = 0:4), that is 1 mm behind the screen. This position is determined so that, when training the ~~xcsf~~ XCSF controller to reproduce the movements generated by the

~~nops~~NOPS, training is performed beyond the screen so as to prevent the effect of discontinuities in the approximated control functions.

When using CEPS, reaching the goal point is replaced by hitting the target on the screen. The target is defined as an interval of varying length around (x = 0; y = 0:399). The movement is stopped once the line y = 0:4 has been crossed. ~~, and t~~The intersect between the trajectory and this line is computed to determine whether the target was hit. The reward for immediately hitting the target without taking the incurred costs into account is set to 40.

In order to train XCSF~~xcsf~~, we generated trajectories with the ~~nops~~ NOPS from each ~~the~~ initial position~~s~~ shown in Fig. 10 to the goal point. The starting points have been organized into five groups of different distances with respect to the goal point (d = 10cm, 12cm, 14cm, 16cm and 18cm respectively). There are 40 initial positions per distance, thus a total of 200 initial positions.

We measure the dispersion over 100 movements towards this target, as well as the average movement times and average movement costs.

We ~~run~~ ran 200 steps of ~~ceps~~ CEPS on the ~~xcsf~~ XCSF controller for each target.

For all the obtained controllers, we measured again the dispersion over 100 movements, the average movement time and average movement cost.

Finally, for all these targets, we recorded the velocity profile under different noise conditions.

A Nomenclature of arm parameters

Table 1. Parameters of the arm model.

m_i	mass of segment i (kg)
l_i	length of segment i (m)
s_i	inertia of segment i ($kg.m^2$)
d_i	distance from the center of segment i to its center of mass (m)
κ	Heaviside filter parameter
A	moment arm matrix ($\in \mathbb{R}^{6\times 2}$)
$f_{\max}$	maximum muscular tension ($\in \mathbb{R}^6$)
M	inertia matrix ($\in \mathbb{R}^{2\times 2}$)
C	Coriolis force ($N.m \in \mathbb{R}^2$)
τ	segments torque ($N.m \in \mathbb{R}^2$)
B	damping term ($N.m \in \mathbb{R}^2$)
u	raw muscular activation (action) ($\in [0,1]^6$)
σ_u^2	multiplicative muscular noise ($\in [0,1]^6$)
$\tilde{u}$	filtered noisy muscular activation ($\in [0,1]^6$)
q^*	target articular position ($rad \in [0,2\pi[^2$)
q	current articular position ($rad \in [0,2\pi[^2$)
$\dot{q}$	current articular speed ($rad.s^{-1}$)
$\ddot{q}$	current articular acceleration ($rad.s^{-2}$)

Legends

Figure 1. Influence of movement time on cost-related quantities~~. Green: subjective utility of hitting the target; red: muscular energy cost; black: global cost versus reward trade-off~~. The red area denotes infeasible short times; blue: probability to hit the target; orange:reward expectation (subjective reward times probability). A: Sketch of the models in [1] and [2]. The subjective utility of hitting the reward (green trace) decreases over time as one is less interested in gains that will occur in a distant future than at the present time. Hitting is less and less costly in terms of efforts (red trace) as the movement is performed more slowly. The expected gain, resulting from the sum of the subjective reward and the (negative) cost reaches a maximum for a certain time (black trace). When the gain is negative (outside the useful interval), one should not move. B: Sketch of the presented model. In the case of a larger target, the hitting probability (blue trace) is higher for faster movements (dashed~~solid~~ lines) than for a smaller target (~~dashed~~ solid lines). As a result, the maximum of the reward expectation (orange trace) is shifted towards longer time for smaller targets, and hence, the optimum movement time is also longer for smaller targets. The red area denotes infeasible short times~~; blue: probability to hit the target; orange:reward expectation (subjective reward times probability).~~ .

Figure 2. Cost of reaching movements towards the star at (x = 0; y = 0:4). The movement is stopped when the end-effector crosses the green line at y = 0:4 (see Methods for details). The color of a point in the reachable space illustrates the cost of a reaching movement from that point. The color-cost correspondence is given by the scale on the ~~right~~ right-hand~~-~~ side. As expected, the smaller the distance to the target, the lower the cost. Furthermore, starting from the ~~left~~ left-hand~~-~~ side of the goal point (A) results in a lower cost than starting from the ~~right~~ right-hand~~-~~ side (B).

Figure 3. Muscular activations when performing a reaching movement, either starting from point A in Figure 2 (left) or from point B in the same Figure (right). The numbers in the legend correspond to the muscles numbers in Figure 9(b). One can see that the pattern of activation is very different depending on the initial configuration of the movement, corresponding to two different optimal strategies: when starting from A, only the shoulder is moved, whereas when starting from B, both the elbow and shoulder are moved, resulting in a more complex muscular activation strategy. One can also see that co-contraction is avoided, consistently with the minimum intervention principle.

Figure 4. A: Time of movement for various target sizes and distances. The larger the target, the faster the movement. Additionnally, the further the target, the longer the movement. B: Expected movement gain for various target sizes and distances. The larger the target, the higher the gain. Additionnally, the further the target, the lower the gain.

Figure 5. Dispersion resulting from the CEPS controllers optimized for five different target sizes, a movement distance of 18cm and a noise amplitude of 0.4. One can observe that the controller often misses small targets whereas it does not use all the potential dispersion for large targets.

Figure 6. Reproduction of Fitts law based on the results of Section 2.2.

Figure 7. Velocity profiles for different amplitudes of noise.

Figure 8. Schematic view of the Cross-Entropy method. A: Start with the normal distribution (mu;sigma2). B: Draw sample parameters from this distribution, evaluate them and select the best ones (in grey). C: Compute the new and sigma2 (adding some noise) and go to A.

Figure 9. Arm model. (a) Schematic view of the arm mechanics. (b) Schematic view of the muscular actuation of the arm, where each number represents a muscle whose name is in the box.

Figure 10. The arm workspace. The reachable space is delimited by a spiral-shaped envelope. The two segments of the arm are represented by two green lines. Initial movement positions are represented with blue dots organized into five sets of different distances to the target. The screen is represented as a dark green line positioned at y = 0~~:~~.4. The origin of the arm is at x = 0:0; y = 0:0.