Optimal Learning Gain Selection in Model Reference Iterative Learning Control Algorithms for Computational Human Motor Systems

Shou-Han Zhou, Denny Oetomo, Ying Tan, Etienne Burdet, and Iven Mareels

Abstract—The role of learning gains in the ability of a computational framework to better capture the behaviour of human motor control in learning and executing a task is the subject of discussion in this paper. In our previous work, a computational model for human motor learning of a task through repetition was established and its convergence analysed. In this paper, the performance of the model is investigated through the addition of degrees of freedom in selecting learning gains, specifically the ability to independently select the learning gain for the damping term. A particle swarm optimisation (PSO) algorithm is utilised to obtain a set of gains optimised to reduce the discrepancy between the experimental data and the simulated trajectories. It is found that it is possible to improve the accuracy of the computational model through the appropriate choice of learning gains. The results and interesting findings are presented and discussed in this paper.

1. INTRODUCTION

In the literature of human motor control and human motor learning, computational models play an ever increasing role. It has been proposed in the literature that two of the main mechanisms which contribute to human motor learning are the formation of an internal model [1][2] and the adjustment of body impedance [3]. Both mechanisms seek to improve the execution of a given tasks over repetitive practice through observations of past attempts. Consequently, in order to learn, it is necessary for the human to behave as a system which performs discontinuous operations repetitively. Control of such systems so as to improve the transient responses and tracking performance has been widely studied in the field of iterative learning control (ILC).

Although various computational models using ILC have been successful in generating results comparable to those measured in the experiment [4], [5], [6], the controllers involve an adjustment of the system inputs states. In computational models of human motor learning, the input states are usually the desired joint acceleration or the desired joint/muscle torques. The adjustment of these states are in contrast to one of the most widely used theories in the motor control literature, which argues that the central nervous system controls the human body through the adjustment of parameters of the motor system such as impedance or damping [7].

Recent models have been proposed to explain how task execution is given effect and improved over time [8], [9], [10], i.e. how human learns to execute a particular task. The proposed computational human motor model identifies the role of ILC in the human motor system. In the computational model developed in [9], a model reference adaptive control strategy combined with iterative learning control (ILC) algorithms was utilised to represent the motor control and adaptation of a person learning to perform a new task through repetitions. The proposed framework was then tuned and validated against a set of experimental data of human subjects performing the task of reaching under dynamic disturbances [3]. The results showed that the proposed computational model based on model reference iterative learning control can well present the human being’s ability to learn unknown environment. The simulation results well match the experimental results. For example, the well-known stages in human learning such as before learning behaviors, after learning behaviors can be clearly observed in the simulation results. This shows the effectiveness of the proposed computational framework.

In the model reference iterative learning algorithms, a learning gain has been selected for each component (unknown time-varying parameter). Learning with respect to position and velocity is treated equally. The convergence of the proposed learning algorithms has been shown in [8], [10]. However, experiments/simulation results show that human beings learn faster than our computational human model does. This suggests that the learning gain selected by trial and error may not be the best.

In addition to convergence (indicating that the task is learned), the rate at which human can learn a new task is quite crucial. As pointed out in [11], stroke patients did not have a learning deficit, but they have weakness-related slowness to develop the required force to implement anticipatory control. The convergence of learning algorithms is dependent on the learning gains. Therefore, the learning gain can potentially be used as a qualitative measure to determine the human being’s ability to learn motor skills.

This paper explores the learning gains of the model reference learning algorithms in the computational model. In other words, it is desired to find the best possible learning gains that produce simulation results closest to that obtained in the experiment. This is done by allowing a set of gains to be determined through an optimization technique and analyzing the resulting behavior. The resulting information may provide insights into the learning and motor control mechanisms that the body / central nervous system (CNS) relies on in carrying out a manipulation task.

Moreover, the computational framework in this work also
investigates the effects of the additional degrees of freedom created when the learning gain weights for position errors and velocity errors are different. By introducing this additional freedom, the search region for the optimization algorithm increases, allowing a better performance in matching the results of the computational model to the experimental results. A Particle Swarm Optimisation (PSO) algorithm [12] was used to provide a quick view on the possibilities of optimising learning gains for a particular task. The discussion in this paper highlights the roles of the learning gains in model reference iterative learning control. The obtained computational model can thus better characterize the ability of human to learn.

II. CONSTRUCTION OF THE MODEL

This section briefly summarizes the computational model developed in [9], [10].

The structure of the proposed framework established in [8][9][10] is shown in Figure 1, based on the schemes of iterative learning control and model reference adaptive control (MRAC).

![Fig. 1: Model Reference Adaptive Controller Framework](image)

Under the scheme, control and learning are achieved through the adjustment of the controller gains, which is in contrast to adjustment of control inputs as used by traditional iterative learning algorithms. The framework consists of three main components:

1) The reference model
2) The plant model
3) The iterative learning controller

The reference model $G_M$ is used to determine the desired output $y_d \in \mathbb{R}^m$ from a given reference input $r \in \mathbb{R}^m$. In this case, our desired output is the ideal motion of our body (e.g. hand) while performing the task set by the reference input. The ideal model can be thought of as the motion of a person performing the task in the absence of any disturbances. Through the process of system identification performed on the experimental data, the reference model was found to be a second order system.

The plant model $G_P$ represents the dynamics of the human body and the unknown, finite environmental disturbances. In the framework, the plant model is formed by constructing a model of the body, such as the upper limb, while the unknown environmental disturbance $d(t)$, is represented by a time-varying, but iteration-invariant disturbances. Under the formulation, the model assumes that the descending motor commands from the CNS regulate the human motor system at the acceleration level in task space. The model of the dynamics of the human body is feedback linearised, to produce an overall plant model expressed as:

$$\dot{y}(t) = Ay(t) + Bu(t) + d(t) \quad \forall t \in [0, T] \quad (1)$$

where $T$ is the time required to achieve the task. Moreover, the disturbance is assumed to satisfy the condition $\max_{t\in[0,T]} \|d(t)\| \leq b_d$, where $b_d$ is an unknown positive constant. More details on the reference model and the plant model can be found in [8][9][10].

The controller is an iterative learning controller which adjusts the controller gain $K$ across each iteration $i$. This is the focus of the study reported in this paper and hence it is presented in a more detailed manner. During each iteration, the controller generates the plant input $u_i$ using the state feedback control law

$$u_i(t) = K_i(t)y_i(t) + K_{2i}(t)r_i(t) - K_{3i}(t)$$

$$= K_i(t)\phi_1(t), \quad \forall t \in [0, T] \quad (2)$$

where $K_i = \begin{bmatrix} K_{1i} & K_{2i} & -K_{3i} \end{bmatrix} \in \mathbb{R}^{2 \times 6}$ and $\phi_i = \begin{bmatrix} \phi_{1i} & \phi_{2i} & \phi_{3i} \end{bmatrix} \in \mathbb{R}^{6 \times 1}$ with $\phi_{1i} = y_i$, $\phi_{2i} = r$ and $\phi_{3i} = \begin{bmatrix} 1 \ 1 \end{bmatrix}^T$.

By denoting the error between the desired and actual outputs at any time instant within one iteration as $e(t) = y_d(t) - y_i(t)$ and assuming that the plant and the reference models satisfy the following conditions:

- The reference model is minimum phase with a known relative degree;
- The plant model is approximated as a stable linear system of order $n$;
- The relative degree of the plant model is finite and equals that of the reference model;
- The order of the reference model ($n_M$) is known and satisfies $n_M \leq n$;
- At each iteration, the initial value satisfies $y_{d}(0) = y_i(0)$,

then the learning algorithm can be expressed as:

$$K_{k,i}(t) = K_{k-1,i}(t) - \beta_k e_i \phi_{k,i}^T, \quad \forall t \in [0, T] \quad (3)$$

where $K_{k,0}(t) = \{0\}_{2 \times 2}$, $k$ is the index for the chosen states and $T$ denoted the transpose. $\beta_k > 0$ is the learning rate for each controller gain. The convergence properties of the proposed learning model reference adaptive controller (3) is shown in [8]. The convergence analysis is based on the composite energy functions proposed in [13]. The learning gain $\beta_k$, $k = 1, 2, 3$ is selected by trial and errors.

Although convergence is guaranteed when a selected learning gain $\beta = \begin{bmatrix} \beta_1 & \beta_2 & \beta_2 \end{bmatrix}^T$ is used, the performance may not be optimal. Experiments show that human beings learn faster than our computational human model. This suggests that the learning gain selected by trial and error
may not be the best. On the other hand, the error vector in (3) is composed of the error in position $e_p$, and the error in velocity $e_v$, $e = \begin{bmatrix} e_p \\ e_v \end{bmatrix}$. In the sequel, Equation (3) is rewritten as

$$K_{k,i}(t) = K_{k,i-1}(t) - \begin{bmatrix} \beta_{k,1} e_{p_i}(t) (\phi_{k,i}(t))^T \\ \beta_{k,2} e_{v_i}(t) (\phi_{k,i}(t))^T \end{bmatrix}$$  \hspace{1cm} (4)

It should be noted that the updating law (3) is a special case of (4) when $\beta_{k,1} = \beta_{k,2} = \beta$. It is intuitively clear that human beings have different sensitivity to position and velocity. Introducing these two gains can well-present this kind of sensitivity difference. On the other hand, optimization algorithms are used to find the optimal gains. The new freedom introduced by two gains will enlarge the search domain and may lead to the better performance.

Indeed, there exists evidence in the literature suggesting that humans adjust parameters of their body system through observation of both position and velocity errors [14]. The ILC (4) adjusts the gain matrices using the errors between the desired and actual position and velocity. By slightly modifying the convergence proof in [9], the error is convergent even in the presence of finite external disturbance. Other than convergence, other performance indices are also needed to evaluate the performance of the proposed computational human motor models. In particular, the work in this paper tries to find the best set of learning gains to represent the learning ability of human beings.

In modelling human motor control, the performance of the algorithm is determined by the error between the simulated trajectory and the trajectory observed in the experiment. The performance of the ideal ILC will result in a simulation trajectory and the trajectory observed in the experiment. The discrepancies is the difference between the desired and actual position and velocity information is different. One question arises being. Moreover, the learning speed for position information is considered with a additional tuning parameter ($\beta$). It is intuitively clear that human beings have different sensitivity to position and velocity. Introducing these two gains can well-present this kind of sensitivity difference. On the other hand, optimization algorithms are used to find the optimal gains. The new freedom introduced by two gains will enlarge the search domain and may lead to the better performance.

More precisely, in order to compare the learning performance, the following performance measure is proposed,

$$e_i \equiv \max_{t \in [0, T]} |e_{i,sim} - e_{i,exp}| \quad \forall i \in [0, N]$$  \hspace{1cm} (5)

where $N$ is the maximum number of trials used in the experiment. Here the discrepancies is the difference between the error of the simulation from the ideal trajectory in the $x$ and $y$ direction during each iteration $e_{i,sim}$ and the error of the experiment from the ideal trajectory in the $x$ and $y$ direction during each attempt $e_{i,exp}$. The error $e_i$ is plotted against the number of iteration as shown in Figure 3. From the figure, it is evident that the errors committed with a additional tuning parameter ($\beta_1, \beta_2$) is significantly lower than the case when only $\beta$ is used. This suggests that there exists appropriate learning gain for human being. Moreover, the learning speed for position information and velocity information is different. One question arises naturally, is it possible to find tuning parameters such that the obtained computational model of human learning system best matches experimental results? The answer to this question is important to qualitatively determine the human being’s ability to learn. Consequently, the learning gains might be used as performance indices to indicate discrepancies (such as impairment) in the motor learning functionalities of a human subject, for example in stroke patients. In this work, particle swarm optimisation algorithms (PSO) are used to find the optimal learning gains that best fit the experimental

![Fig. 2: Human subject hand motion during the learning process of a task of reaching within a viscous disturbance field that pushes the hand to the left. (a) Experimentally obtained results (b) simulated with a set of user selected (fixed) learning gain (c) simulated with a $\beta_1$ and $\beta_2$ of learning gain obtained through optimisation.](image)
data (Section III). This would specifically address the roles of the learning gains within the computational model proposed in this paper. The implications of these roles within human motor learning systems is to be further analysed in future work.

III. Optimising Learning Gains

In the computational model, each state considered in the controller (2) is associated with a gain matrix $K$. The iterative learning algorithm (4) adjusts the gain matrices using the two errors $e_p$ and $e_v$.

Early work has assumed equal contributions of these errors in the convergence of the controller gains $K$ towards a steady-state value in the iteration domain. In this paper, this assumption is removed and $\beta_1$ and $\beta_2$, which govern the contribution of each error to the convergence of $K$, are allowed to vary independently. As such, each contribution in (4) is given the freedom to converge to its respective optimum values at different rates, resulting in a separation of time scales for the respective error dynamics. Following this concept, there exists the possibility of an ideal set of learning gains $\beta^*$ to be used by the computational model such that the trajectory simulated follows that of the trajectories observed in the experiment.

From the definition of the controller as shown in (2), it is established that the parameter space, in which the ideal set of learning gains exists, is six dimensional. Particle swarm optimisation (PSO) algorithm [12] is used to obtain an ideal set of learning gains. PSO is a computational method which optimises a given measure of quality by attempting to improve candidate solutions of the computation model. This does not guarantee a global optimal solution, however, it serves the purpose of the study in this paper and is a highly practical method given the size of the problem.

The measure of quality is represented in this paper by a cost function $J$ which is constructed as the total discrepancies between the simulated trajectory and the experiment trajectory at each time instant of one iteration and across all the trials performed. The ideal trajectory is the optimum trajectory made for the given task. The cost function therefore is given as:

$$J = \int_0^T \int_0^N (\xi_i(t))^T (\xi_i(t)) dt di, \quad \forall t \in [0, T]$$ (6)

where $\xi_i(t) = e_i^{\text{sim}}(t) - e_i^{\text{exp}}(t)$ represents the error between the simulation and experiment for the two directions. The errors are calculated using the data sets obtained from the experiment and generated from the simulation, both of which are described in the following sections.

IV. Experimental Data Set

In the experiment, a human subject was to perform the task of reaching for a target location while holding the robot end-effector. The task is planar and the target is located directly in front of the human subject. Disturbances are generated by the robot attached to the human arm and the objective is for the human subject to learn to perform the task of reaching for the target in the face of the injected disturbances. Two studies were conducted in which the robot generates disturbances dependent on the hand end-effector velocity (Velocity Field) and end-effector position (Divergent Field). Details of the experimental setups are given in [4] [5] [10].

**Velocity Field** or VF. The robot generated a disturbance force field dependent on the subject’s end-effector velocity in the cartesian coordinates, as described by:

$$\begin{bmatrix} D_{vx} \\ D_{vy} \end{bmatrix} = -\begin{bmatrix} 13 & -18 \\ 18 & 13 \end{bmatrix} \begin{bmatrix} P_x \\ P_y \end{bmatrix}$$ (7)

where $D_v = [D_{vx} \ D_{vy}]^T \in \mathbb{R}^2$ was the disturbance forces applied to the hand/robot end-effector in the Cartesian coordinates and $P = [P_x \ P_y]^T \in \mathbb{R}^2$ was the position of the hand end-effector.

**Divergent Field** or DF. In this case, the robot generated a Divergent Field (DF) environment, which was dependent on the subject’s end-effector position in the generalized cartesian coordinates, as described by the following equation:

$$\begin{bmatrix} D_{dx} \\ D_{dy} \end{bmatrix} = -\begin{bmatrix} 450 & 0 \\ 0 & 450 \end{bmatrix} \begin{bmatrix} P_x \\ P_y \end{bmatrix}$$ (8)

where $D_d \in \mathbb{R}^2$ was the disturbance forces applied to the hand/robot end-effector in cartesian coordinates. The purpose of this artificial environment was to observe evidences of impedance adjustment within the human motor system during learning. The study in the DF environment is conducted in the same manner as that within the VF environment described in Case 2.

Throughout this paper, the hand end-effector position data for the two experiments are presented for one subject for the purpose of clear illustration. The set of data presented is representative of the overall behaviour observed in this
objective, and hence can be used to describe the characteristics of the human motor system without significant loss of generality.

V. SIMULATING HUMAN MOTOR LEARNING ON THE PROPOSED COMPUTATIONAL FRAMEWORK

The optimisation algorithm from Section III is incorporated into the computational framework described in Section II. The framework is used to simulate the behaviour of human motor learning in a task where human subject is required to reach for a target location while subject to the environmental disturbances used in the experiment. The optimisation technique is carried out by adjusting the learning gains, $\beta_1$ and $\beta_2$ in order to obtain the simulated results that best mimic the learning behaviour as demonstrated by the experimental data set:

$$\beta^* = \left(\frac{\beta_1^*}{\beta_2^*}\right) = \arg \min_{\beta} (J)$$

(9)

where $J$ is defined in (6).

The following assumptions are made in the simulation:

1) Independent control systems are used to control motion in the generalised $x$ and $y$ directions in task space. This follows from the experimental evidence suggesting that for the task of planar reaching, there exist two sets of control such that the human motion in task space is decoupled along their respective generalised coordinates [15].

2) The inertia of the robot is assumed to be negligible, hence its dynamics is not considered in the simulation model. This is justified since the robot used for the experiment was designed to produce very little apparent inertia through good balancing and also because the subjects were allowed to familiarise themselves with the dynamic of the robot before the actual experiment trials.

3) Noise is modelled as a band limited Gaussian white noise to the reference input $r$. This implies that for a given movement, only a crude ideal trajectory is required to be planned by the CNS for a given task [16].

In order to demonstrate the changes in the performance of the ILC algorithm, the computational model is simulated for following settings:

$$\beta_1 = \left[ \begin{array}{ccc} \beta_{1,1} & \beta_{2,1} & \beta_{3,1} \end{array} \right]^T$$

$$\beta_2 = \left[ \begin{array}{ccc} \beta_{1,2} & \beta_{2,2} & \beta_{3,2} \end{array} \right]^T$$

- **Case 1**: The set of learning gains $\beta_1 = \beta_2 = \beta = \left[ \begin{array}{ccc} 300 & 300 & 10 \end{array} \right]^T$ is used. These values are obtained from previous work [8].

- **Case 2**: The learning gains are unconstrained and optimised through the particle swarm optimisation technique. While the technique does not guarantee global optimality, the resulting behaviour has been through a systematic search in the solution space and provide a set of learning gains that minimises the cost function (6).

These two cases are simulated for the Velocity Field (VF) and Divergent Field (DF) environments. The results are presented and discussed in the following section.

VI. RESULTS AND DISCUSSION

Recently, human motor learning has been modeled using a computational model based on model reference adaptive control and iterative learning control. Although stability and convergence of the controller has been rigorously demonstrated [9], less attention is focused on the performance of the algorithm. In this paper, the performance of the controller, which is governed by the learning gains of the algorithm $\beta$, is analysed. In previous models, the learning gains were assumed to be constrained. This assumption is relaxed in this paper in order to use the learning gains as a design variable to optimise the performance of the iterative learning controller.

The results of both cases (Case 1 and Case 2) along with the experiment results are given in Figure 4 and Figure 5 for the velocity field and the divergent field environments, respectively.

The results are decomposed into five phases, representing the stages in which learning occurs. These phases have been referred previously in the literature [3][17]. In analysing each phase, it is evident that for both environments, both cases were able to generate similar learning trends observed in the experiment. Nevertheless, it is also observed that the trajectories simulated for each case are significantly different for each phases of learning. Moreover, it seems that unconstrained learning gains has the potential to significantly improve the simulation performance for specific phases.

Taking the the velocity field as an example, it is evident that during the before learning phase, in which the subjects are first exposed to the new environment, Case 1 simulation is able to capture the experiment observation that the subject’s end-effector only approaches the target and is unable to reach it within the specified time. Case 2 generated trajectories where initial maximum deviation from the straight line are more similar to those of the experiment. In observing the after learning phase of the velocity field, it is evident that the trajectories of Case 2 are significantly closer to that of the experiment compared to that of Case 1.

The optimised $\beta_1$ is identified to be significantly larger in magnitude than that $\beta_2$. The values of the learning gains for Case 2 after optimisation is given as:

$$\beta_1^* = \left[ \begin{array}{ccc} 105 & 123 & 86 \end{array} \right]^T$$

$$\beta_2^* = \left[ \begin{array}{ccc} 65 & 53 & 6 \end{array} \right]^T$$

Considering that $\beta_1$ and $\beta_2$ represent the contributions of the position and velocity errors respectively, this implies that the CNS prioritise $e_p$ over $e_v$ during learning. As a better fit to the experimental data is obtained through the use of different learning gain for $\beta_1$ and $\beta_2$, it can also be
implied that humans utilise the degree of freedom to isolate and learn the damping component of its hand motion during the execution of a new task. Much of the current research has been focused on the roles of impedance and the use of internal model. These implications will be subject to future research on real subjects.

Another interesting observation is that in the velocity field, during the before learning phase, the simulation in Case 1 is closer to that of the experiment than that of Case 2 while during the after learning phase, the simulation of case 2 is closer. Therefore, the controller may require different sets of update rates before and after learning in order to fully capture the experiment behaviour. This means that the CNS prioritises different states to adjust the body parameters during learning, implying that the learning gains $\beta_1$ and $\beta_2$ may be required to be adjusted. This hypotheses agrees with the findings of [18] and may be subject to future work.

VII. CONCLUSION

In this paper, the role of learning rates/gains is investigated for a recently proposed computational model of human motor learning. From the iterative learning algorithm, the learning gains are identified to govern the convergence of the controller and therefore can affect the performance of the simulation in matching the experiment behaviours. In this paper, the gains are optimised in order to seek for the best gains so as to minimise the discrepancies between the simulation and the experiment.

The results of this paper suggests that although, different learning gains are superior in capturing a specific feature of the experiment, there exist no optimal learning gains exist which can fully capture all the behaviour. The implications of this result is that the CNS may prioritise the states it uses to learn and, in addition, the priorities of the states are adjusted during learning.

REFERENCES


Divergent Field

Experiment Results

Simulation: Case 1

Simulation: Case 2

Fig. 5: Comparison of the experimentally measured response and the simulation results produced using the proposed computational model in Divergent Field (DF) environment


