Bayesian Modelling of a Sensorimotor Loop: Application to Handwriting

Estelle Gilet, Julien Diard, Pierre Bessiere

Bayesian Modelling of a Sensorimotor Loop: Application to Handwriting

Estelle GILET(1) **Julien Diard**(2), **Pierre Bessière**(1)

(1)Laboratoire d'Informatique de Grenoble
CNRS – INRIA Rhône-Alpes
655, avenue de l'Europe
38330 Montbonnot Saint Martin
estelle.gilet@inrialpes.fr

(2)Laboratoire de Psychologie et NeuroCognition
CNRS – Université Pierre Mendès France
BSHM – BP47
38040 Grenoble Cedex 9

Abstract

This paper concerns the Bayesian modelling of a sensorimotor loop. We present a preliminary model of handwriting, that provides both production of letters and their recognition. It is structured around an abstract internal representation of letters, which acts as a pivot between motor and sensors models. The representation of letters is independent of the effector usually used to perform the movement. We show how our model allows to solve a variety of tasks, like letter reading, recognizing the writer, and letter writing (with different effectors). We show how the joint modelling of the sensory and motor systems allows to solve reading tasks in the case of noisy inputs by internal simulation of movements.

1 Introduction

If you were asked to write down your name, you would probably consider it a mundane task. You could surely perform it easily in a variety of circumstances, like thinking about something else, looking elsewhere, etc. But what about writing your name with your foot, in the sand or snow, for instance? It turns out that this, too, is rather easy. The performed trace would be somewhat distorted from your handwriting, but, even without any training in "footwriting", your name would be readable.

This effect is known as *motor equivalence* [9]. It has been used as an evidence that internal representations of movements might be independent of the effector usually used to perform them. This idea has been used in mathematical models of both movement production and recognition.

Models of movement production (minimum-jerk [1], energy [5], torque change [8], variance [2]) and systems of handwriting recognition usually only describe one half of the problem, either the production side, or the recognition side. For instance, a purely motor model would not be able to both produce letters and care about their readability. On the other hand, a purely sensory model, for instance performing Optical Character Recognition (OCR) based on Independent or Principal Component Analysis (ICA, PCA), would internally describe letters in some image-based space which might not be easily used for producing handwriting. In other words, these approaches are hemiplegic in nature: they either define systems expert at reading, but that could not sign their names, or expert at writing, but that are blind.

If handwriting is a perception-action cycle, it probably is fruitful to consider it as a whole. Some modeling frameworks have already been proposed to study the interplay between perception and action in sensorimotor processes, like the motor theories of perception [4] or the perception for action control theory [7]. However, they are mostly conceptual models, and lack mathematical implementations. The Bayesian or subjective probabilistic formalism is, in this context, a suitable

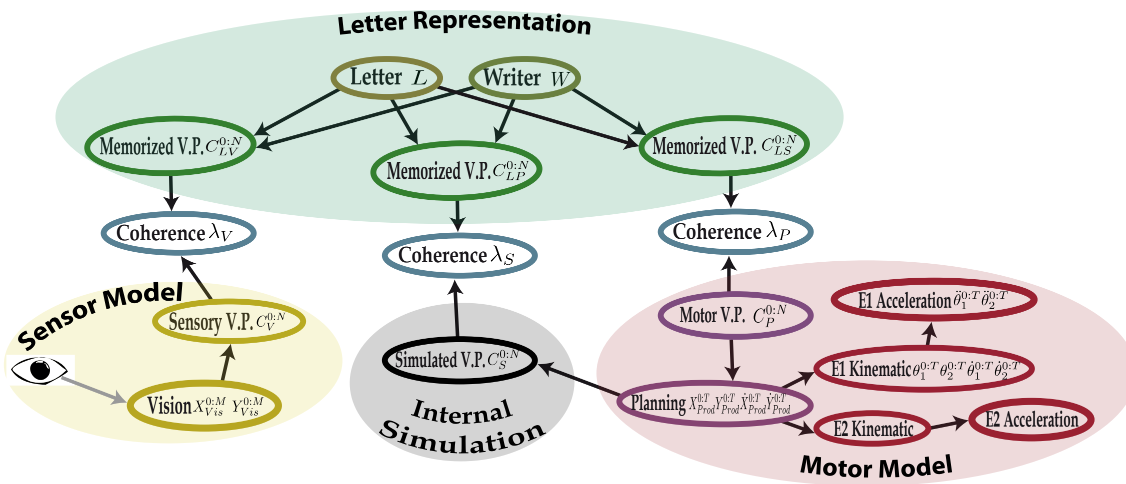

Figure 1: Global structure of the model: the representation of letters is the pivot between the motor and sensor models. V.P. stands for via-points ; $E1$ and $E2$ refer to different effectors (*e.g.* an arm, a leg).

tool. It is based on a unique mathematical framework, which is used to model both the articulation between parts of the model, and the parts themselves.

We therefore propose, in this paper, a preliminary Bayesian model of the sensorimotor loop involved in handwriting. The model provides both production of letters and their recognition.

It is structured around an abstract internal representation of letters, which acts as a pivot between motor and sensor models (Figure 1). Letters are represented internally by sequences of via-points, which are distinctive points along the trajectory. The motor model is made of two parts, related respectively to the planification of the trajectory and to the effector description. In this preliminary work, we only include a very simplified, high-level vision model, that extracts geometric properties of the trajectory. Finally, an internal simulation model allows to predict the effects of motor movements.

We define a complete Bayesian model, which articulates these four components: abstract representation of letters, motor model, sensor model and internal simulation (Section 2). It allows to solve a variety of cognitive tasks, from writing (with different effectors) to reading (reading complete letters, reading letters as they are being traced, recognizing the writer, reading letters in difficult cases). Each of these is defined mathematically by a probabilistic question to the global model, and is solved automatically by Bayesian inference (Section 3).

2 Model

We give here the formal definition of the joint probability distribution of the global model. It just defines the articulation between sub-models. The representation of letters is the pivot between sensor models and motor models (Figure 1).

$$
\begin{aligned}
& P\left(\begin{array}{l} C_{LV}^{0:N}\, C_{LP}^{0:N}\, C_{LS}^{0:N}\, C_{V}^{0:N}\, \lambda_P\, \lambda_S\, \lambda_V\, \theta_1^{0:T}\, \theta_2^{0:T}\, \dot{\theta}_1^{0:T}\, \dot{\theta}_2^{0:T}\, \ddot{\theta}_1^{0:T}\, \ddot{\theta}_2^{0:T} \\ C_P^{0:N}\, C_S^{0:N}\, X_{Vis}^{0:M}\, Y_{Vis}^{0:M}\, X_{Prod}^{0:T}\, Y_{Prod}^{0:T}\, \dot{X}_{Prod}^{0:T}\, \dot{Y}_{Prod}^{0:T}\, L\, W \end{array}\right) \\
&= \left(\begin{array}{l} P(C_{LV}^{0:N} \mid L\, W)\, P(X_{Vis}^{0:M}\, Y_{Vis}^{0:M})\, P(\lambda_V \mid C_{LV}^{0:N}\, C_V^{0:N})\, P(C_{LP}^{0:N} \mid L\, W)\, P(\lambda_P \mid C_{LP}^{0:N}\, C_P^{0:N}) \\ P(C_P^{0:N})\, P(X_{Prod}^{0:T}\, Y_{Prod}^{0:T}\, \dot{X}_{Prod}^{0:T}\, \dot{Y}_{Prod}^{0:T} \mid C_P^{0:N})\, P(\ddot{\theta}_1^{0:T}\, \ddot{\theta}_2^{0:T} \mid \theta_1^{0:T}\, \theta_2^{0:T}\, \dot{\theta}_1^{0:T}\, \dot{\theta}_2^{0:T}) \\ P(\theta_1^{0:T}\, \theta_2^{0:T}\, \dot{\theta}_1^{0:T}\, \dot{\theta}_2^{0:T} \mid X_{Prod}^{0:T}\, Y_{Prod}^{0:T}\, \dot{X}_{Prod}^{0:T}\, \dot{Y}_{Prod}^{0:T})\, P(C_S^{0:N} \mid X_{Prod}^{0:T}\, Y_{Prod}^{0:T}\, \dot{X}_{Prod}^{0:T}\, \dot{Y}_{Prod}^{0:T}) \\ P(C_{LS}^{0:N} \mid L\, W)\, P(\lambda_S \mid C_{LS}^{0:N}\, C_S^{0:N})\, P(C_V^{0:N} \mid X_{Vis}^{0:M}\, Y_{Vis}^{0:M})\, P(L)\, P(W) \end{array}\right) \quad (1)
\end{aligned}
$$

We now describe each of these sub-models in turn, providing the details about variables and their meanings.

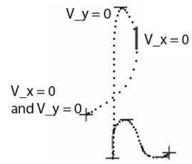

Figure 2: Representation of letters with via-points.

2.1 Representation of letters

We assume that a letter is internally represented by a sequence of via-points, that are part of the whole X_{Vis}, Y_{Vis} trajectory of the letter. We further assume that these points are also encoded in the allocentric reference frame, as opposed to the articulatory reference frame.

We restrict via-points to places in the trajectory where either the X derivative ($\dot{X}$) or the Y derivative ($\dot{Y}$), or both, is zero (Figure 2). When this occurs, this creates a salient point, both from a motor perspective, as this means the movement changes direction, and from a sensory perspective, as this means the trajectory is at a local extremum.

For each via-point (C_{LV}), each writer and each letter, we encode Gaussian probability distributions over its 2D position (X_{LV}, Y_{LV}), and the velocity of passage at this via-point ($\dot{X}_{LV}, \dot{Y}_{LV}$) (one of which is a sharp distribution centered on 0, by definition of via-points).

The set of via-points is denoted as $C_{LV}^{0:N}$: $C_{LV}^n = \{X_{LV}^n, Y_{LV}^n, \dot{X}_{LV}^n, \dot{Y}_{LV}^n\}$. n is used as an index in the sequence of via-points. The joint distribution over this set of variables is defined as:

$$P(C_{LV}^{0:N} \mid L\, W) = \prod_{n=0}^{N} P(X_{LV}^n \mid L\, W)\, P(Y_{LV}^n \mid L\, W)\, P(\dot{X}_{LV}^n \mid L\, W)\, P(\dot{Y}_{LV}^n \mid L\, W). \qquad (2)$$

The product of terms indicates that via-points are considered independent of each other if the letter and writer are known.

For algorithmic reasons, the term $P(C_{LV}^{0:N} \mid L\, W)$ is present three times in the decomposition of the joint probability distribution, one for each sub-part of the model: $P(C_{LV}^{0:N} \mid L\, W)$, $P(C_{LP}^{0:N} \mid L\, W)$ and $P(C_{LS}^{0:N} \mid L\, W)$. They just differ in the name of variables.

2.2 Vision model

We assume a simple vision model, that concerns the extraction of via-points from trajectories, using their geometric properties. The term $P(C_V^{0:N} \mid X_{Vis}^{0:M}\, Y_{Vis}^{0:M})$ describes how the via-points are extracted from a trajectory. This follows from our via-point definition: when $\dot{X}$ or $\dot{Y}$ is null, then a new via-point is found and the position and velocity profiles are encoded. We define this term by Dirac probability distributions.

2.3 Motor model

The motor model is made of two parts: the trajectory formation and the effector control. We note that the trajectory formation is independant of the effector generally used to perform the movement.

2.3.1 Trajectory formation

The term $P(X_{Prod}^{0:T}\, Y_{Prod}^{0:T}\, \dot{X}_{Prod}^{0:T}\, \dot{Y}_{Prod}^{0:T} \mid C_P^{0:N})$ concerns general trajectory formation. It is expressed in the Cartesian reference frame, and is defined by Dirac probability distributions:

$$\begin{aligned} & P(X_{Prod}^{0:T}\, Y_{Prod}^{0:T}\, \dot{X}_{Prod}^{0:T}\, \dot{Y}_{Prod}^{0:T} \mid C_P^{0:N}) \\ = & \prod_{n=0}^{N-1} P(X_{Prod}^{Kn:K(n+1)}\, Y_{Prod}^{Kn:K(n+1)}\, \dot{X}_{Prod}^{Kn:K(n+1)}\, \dot{Y}_{Prod}^{Kn:K(n+1)} \mid C_P^n\, C_P^{n+1}) \text{ with } KN = T \end{aligned}$$

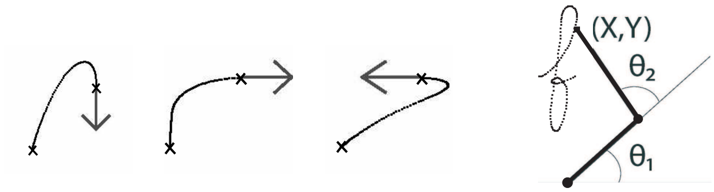

Figure 3: Left: The generation of trajectories between an initial point and a given position to attain. The difference between these three cases is the velocity to attain at the final point (left: $\dot{X} = 0$ and $\dot{Y} < 0$, middle: $\dot{X} > 0$ and $\dot{Y} = 0$, right: $\dot{X} < 0$ and $\dot{Y} = 0$). Right: A two-joint manipulator, the articulatory and end-point position variables.

The term inside the product describes the computation of K intermediary points, between an initial position C_P^n and a given position to attain C_P^{n+1}. A common robotic algorithm helps define this term: an acceleration profile is chosen, that constrains the interpolation. In our case, we used a "bang-bang" profile [3], where the effector first applies a maximum acceleration, followed by a maximum negative acceleration (Figure 3, left). The obtained trajectory is a degree 5 polynomial, that is also the polynomial solution of minimum jerk [1].

2.3.2 Effector model

The effector model is made of two parts, related to the geometry of the considered effector (kinematic model) and the control of this effector for general movement production (acceleration model).

Kinematic model The kinematic model describes the geometry of the effector, and provides direct and inverse transforms between endpoint and articulatory coordinates.

In our simulation, the human arm is represented by a two-joint manipulator (Figure 3, right): θ_1 represents the shoulder angle and θ_2 represents the elbow angle. The endpoint position is described by the cartesian coordinates X and Y.

The term $P(\theta_1^{0:T}\ \theta_2^{0:T}\ \dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T} \mid X_{Prod}^{0:T}\ Y_{Prod}^{0:T}\ \dot{X}_{Prod}^{0:T}\ \dot{Y}_{Prod}^{0:T})$ describes the inverse kinematic transform, which translates the endpoint cartesian coordinates to articulatory angles. We define this term by Dirac probability distribution.

Acceleration model The term $P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid \theta_1^{0:T}\ \theta_2^{0:T}\ \dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T})$ concerns the computation of successive derivatives using finite differences, more precisely the probability distribution over accelerations given the velocities at time t and $t-1$.

2.4 Internal simulation

The simulation model concerns the extraction of via-points from internally simulated trajectories. The term $P(C_S^{0:N} \mid X_{Prod}^{0:T}\ Y_{Prod}^{0:T}\ \dot{X}_{Prod}^{0:T}\ \dot{Y}_{Prod}^{0:T})$ is defined like in the vision model.

3 Using the probabilistic model

We have shown how the components of our global model are defined. Therefore the joint probability distribution (1) is specified and the model is fully defined. Thanks to Bayesian inference, it can be used to automatically solve cognitive tasks. All inferences are carried out by a general purpose probabilistic inference engine (ProBT© of ProBayes). We define a cognitive task by a probabilistic term to be computed, which we call a question. We use the variables λ (coherence variables [6]) to enable ($\lambda = 1$) or disable (summation over λ) some parts of the model.

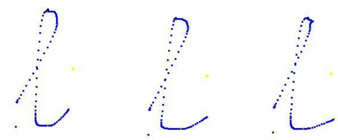

Figure 4: Examples of handwriting production.

3.1 Perception: reading letters, recognizing writers

Given a trajectory $(X_{vis}^{0:M}\ Y_{vis}^{0:M})$, what is the letter? This is solved by computing the following question using Bayesian inference:

$$P(L \mid X_{vis}^{0:M}\ Y_{vis}^{0:M}\ [\lambda_V = 1]) \quad \propto \quad \sum_W \prod_{n=0}^{N} P([C_{LV}^n = C_V^n = f(X_{Vis}^{0:M}, Y_{Vis}^{0:M})] \mid L\ W) \tag{3}$$

This question only involves terms from the representation of letters model and the vision model.

This assumes a fully defined model of the representation of letters: some of its terms have to be learned beforehand. More precisely, $P(X_{LV}^n \mid L\ W)$, $P(Y_{LV}^n \mid L\ W)$, $P(\dot{X}_{LV}^n \mid L\ W)$ and $P(\dot{Y}_{LV}^n \mid L\ W)$ are Gaussian distributions, one for each pair $\langle L, W \rangle$, and which are defined by their means μ and variances σ which are experimentally identified. They were learned in a supervised manner: for each pair $\langle L, W \rangle$, the mean and the variance of the position and velocity of via-points is computed, on a learning data set of 15 examples for each letter.

The model was then tested on a 5*26 test data set: we obtained a high correct recognition rate (89.52%). Misclassifications arise due to the similitude of some letters: l's and e's are similar for the model, probably because the letter size is normalized in the acquisition of data.

Given a trajectory, the model can also recognize the writer, by computing $P(W \mid X_{Vis}^{0:M}\ Y_{Vis}^{0:M}\ [\lambda_V = 1])$. It is solved in a manner similar to (3), except that, instead of summing over W, the summation is over L.

3.2 Action: writing letters

Our model allows to solve the writing task, by computing $P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid L\ W\ [\lambda_P = 1])$.

What are the accelerations to apply to the arm to write a letter? We apply Bayesian inference to answer the question:

$$\begin{array}{l} P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid L\ W\ [\lambda_P = 1]) \\ \propto \left(\begin{array}{l} P(X_{Prod}^{0:T}\ Y_{Prod}^{0:T}\ \dot{X}_{Prod}^{0:T}\ \dot{Y}_{Prod}^{0:T} \mid C_P^{0:N})\ P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid \dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T}\ \theta_1^{0:T}\ \theta_2^{0:T}) \\ P(\dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T}\ \theta_1^{0:T}\ \theta_2^{0:T} \mid X_{Prod}^{0:T}\ Y_{Prod}^{0:T}\ \dot{X}_{Prod}^{0:T}\ \dot{Y}_{Prod}^{0:T}) \\ \sum_{C_{LP}^{0:N}} P([\lambda_P = 1] \mid C_P^{0:N}\ C_{LP}^{0:N})\ P(C_{LP}^{0:N} \mid L\ W) \end{array} \right) . \end{array} \tag{4}$$

We use a two-step algorithm: the representation of letters model is first used to draw the positions and velocities of via-points. Then, the model of trajectory formation determines the trajectory between via-points. The effector model translates these endpoint cartesian coordinates of via-points to articulatory coordinates. Obviously, the vision model is not involved in this question. Figure 4 shows letters obtained in response to the question $P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid [L = l]\ W\ [\lambda_P = 1])$.

3.3 Perception and action: reading letters with internal simulation of writing

When it is difficult to recognize a letter, it could be interesting to internally simulate movements which would produce similar observations: we predict the perceptual outcomes of these simulated actions and match the results with memorized letters. The question corresponding to this task is close to the reading question; we just have to activate the motor model ($\lambda_P = 1$) and the simulation models ($\lambda_S = 1$):

$$P(L \mid X_{vis}^{0:M}\ Y_{vis}^{0:M}\ [\lambda_V = 1]\ [\lambda_P = 1]\ [\lambda_S = 1]) . \tag{5}$$

We apply Bayesian inference to answer the question:

$$
\begin{aligned}
& P(L \mid X_{vis}^{0:M}\ Y_{vis}^{0:M}\ [\lambda_V = 1]\ [\lambda_P = 1]\ [\lambda_S = 1]) \qquad (6) \\
\propto \quad & \sum_W \left(\begin{array}{l} P([C_{LV}^{0:N} = C_V^{0:N} = f(X_{Vis}^{0:M}, Y_{Vis}^{0:M})] \mid L\ W) \\ \sum_{C_{LP}^{0:N}} \left(\begin{array}{l} P(X_{Prod}^{0:T}\ Y_{Prod}^{0:T}\ \dot{X}_{Prod}^{0:T}\ \dot{Y}_{Prod}^{0:T} \mid [C_P^{0:N} = C_{LP}^{0:N}]) \\ P(C_{LS}^{0:N} = C_S^{0:N} = f(X_{Prod}^{0:T}, Y_{Prod}^{0:T}) \mid L\ W) \end{array} \right) \end{array} \right) \\
\propto \quad & \left(\begin{array}{l} P(L \mid X_{vis}^{0:M}\ Y_{vis}^{0:M}\ [\lambda_V = 1]) \\ P(X_{Prod}^{0:T}\ Y_{Prod}^{0:T}\ \dot{X}_{Prod}^{0:T}\ \dot{Y}_{Prod}^{0:T} \mid L\ [\lambda_P = 1]) \\ \sum_W P(C_{LS}^{0:N} = C_S^{0:N} = f(X_{Prod}^{0:T}, Y_{Prod}^{0:T}) \mid L\ W) \end{array} \right) .
\end{aligned}
$$

This question is composed of a product of three terms, two of which are the solutions to previous tasks: the term $P(L \mid X_{vis}^{0:M}\ Y_{vis}^{0:M}\ [\lambda_V = 1])$ amounts to the reading task and $P(X_{Prod}^{0:T}\ Y_{Prod}^{0:T}\ \dot{X}_{Prod}^{0:T}\ \dot{Y}_{Prod}^{0:T} \mid L\ [\lambda_P = 1])$ corresponds to the writing task planning phase. The last term loops the internal simulation in order to compare simulated movements with stored letter representations. Because the three probability distributions are multiplied, their signal is going to be increased; in other words, if the sensory input is ambiguous, the motor simulation will favour letters which look the most like what the system would be able to produce.

4 Conclusion

We have presented a preliminary Bayesian model of handwriting. This model describes the articulation between each sub-models: the representation of letter is the pivot between the motor and sensor models. We have shown preliminary experimental results that highlight the production of letters, their recognition, the writer recognition, and the mathematical inference for reading in difficult cases, using internal simulation of movements.

Current work aims at first experimentally comparing results for these various cognitive tasks, in particular measuring the extent of the benefit of the internal simulation of movements, that is to say, characterise the class of noisy inputs for which the involvement of the motor model is the most useful. We also would like to develop the symmetrical case, “writing with the goal of being readable”, in order to test the limits of our model of the full sensorimotor loop involved in handwriting.

Acknowledgments

This work has been supported by the BACS European project (FP6-IST-027140).

References

[1] T. Flash and N. Hogan. The coordination of arm movements: An experimentally confirmed mathematical model. *Journal of Neuroscience*, 5(7):1688–1703, 1985.

[2] C. Harris and D. Wolpert. Signal-dependent noise determines motor planning. *Nature*, 394:780 – 784, 1998.

[3] W. Khalil and E. Dombre. *Modeling, Identification and Control of Robots*. Hermes Penton Science, 2002.

[4] A. Liberman and I. Mattingly. The motor theory of speech perception revised. *Cognition*, 21:1–36, 1985.

[5] W. Nelson. Physical principles for economies of skilled movements. *Biological Cybernetics*, 46:135 – 147, 1983.

[6] C. Pradalier, F. Colas, and P. Bessière. Expressing bayesian fusion as a product of distributions: Applications in robotics. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems*, 2003.

[7] J.-L. Schwartz, C. Abry, L. Boë, and M. Cathiard. *Phonology in a theory of perception-for-action-control.* Oxford University Press, 2002.

[8] Y. Uno, M. Kawato, and R. Suzuki. Formation and control of optimal trajectory in human multijoint arm movement - minimum torque-change model. *Biological Cybernetics*, 61:89–101, 1989.

[9] A. Wing. Motor control: Mechanisms of motor equivalence in handwriting. *Current Biology*, 10:245–248, 2000.