Bayesian modelling of sensorimotor systems: Application to handwriting

Estelle Gilet, Julien Diard, Pierre Bessiére

BAYESIAN MODELLING OF SENSORIMOTOR SYSTEMS: 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
(2)Laboratoire de Psychologie et NeuroCognition
CNRS – Université Pierre Mendès France
BSHM – BP47
38040 Grenoble Cedex 9

ABSTRACT
This paper concerns Bayesian modeling of a sensorimotor system. We present a preliminary model of handwriting, in which the representation of letter is abstract, and is a pivot between motor and sensor models. We show how our model allows to solve a variety of tasks, like letter reading, recognizing the writer, and letter writing.

KEY WORDS
Bayesian modeling, writing, reading, trajectory planning.

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* [1]. 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 both in mathematical models of movement production and recognition.

Indeed, a large class of models of movement production defines an objective function, which is a measure of performance for possible movements. The performance is measured as the time integral of some cost like jerk (rate of change of acceleration) [2], energy [3], torque change [4], variance [5]. This class of works assume that out of all possible solutions for producing the desired trajectory, the central nervous system selects the one minimizing the chosen measure of performance. In these approaches, when they are applied to handwriting modeling, letters are seen as sequences of points [6] or concatenation of strokes [7].

On the other hand, in handwriting recognition systems, letters are represented using strokes [8], downstrokes, points... These purely sensory models are tailored for good recognition rates, independently of effector models. In other words, the used representations would not necessarily be adequate for writing production.

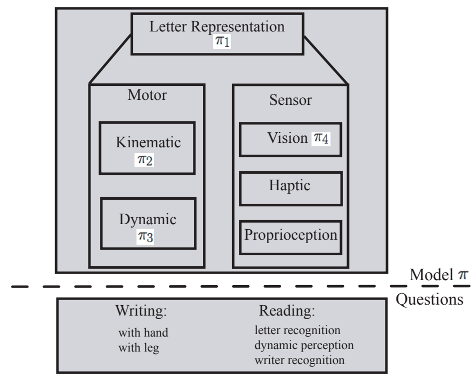

Figure 1. Global structure of the model: the representation of letters is the pivot between the motor and sensor models.

However, both these types of models of handwriting only describe one half of the problem, either the production side, or the recognition side. If handwriting is a sensorimotor process, it probably is fruitful to consider it as a whole. Indeed, modeling frameworks have been proposed, to study the interplay between perception and action in sensorimotor processes, like the motor theories of perception [9] or the perception for action control theory [10].

These are mostly conceptual models, and lack mathematical implementations. The Bayesian or subjective probabilistic formalism is, in this context, a suitable 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 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 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 to the geometry of the considered effector (kinematic model) and the control of this effector for general move-

*Corresponding author: Estelle.Gilet@inrialpes.fr. This work has been supported by the BACS European project (FP6-IST-027140).

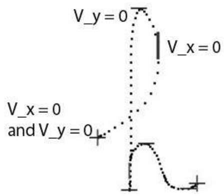

Figure 2. Representation of letters with via-points.

ment production (dynamic model). Letter perception is varied, because letters can be seen, but writing also produces tactile and proprioceptive inputs. In this preliminary work, we only include a very simplified, high-level vision model, that extracts geometric properties of the trajectory.

A complete Bayesian model is mathematically defined, that articulates these three components: abstract representation of letters, motor model, sensor model (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, etc). 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. We note π the global model, and π_i each of the sub-models: π_1 is the representation of letters, π_2 is the kinematic model, π_3 is the dynamic model, and π_4 is the vision model (Figure 1).

$$
\begin{aligned}
& P\left(\begin{array}{l} \theta_1^{0:T}\ \theta_2^{0:T}\ \dot\theta_1^{0:T}\ \dot\theta_2^{0:T}\ \ddot\theta_1^{0:T}\ \ddot\theta_2^{0:T}\ I_p\ Letter\ W \\ X^{0:T}\ Y^{0:T}\ \dot X^{0:T}\ \dot Y^{0:T}\ X_{vp}\ Y_{vp}\ \dot X_{vp}\ \dot Y_{vp} \end{array} \middle| \pi \right) \qquad (1) \\
&= P(X_{vp}\ Y_{vp}\ \dot X_{vp}\ \dot Y_{vp}\ I_p\ Letter\ W \mid \pi_1) \\
&\quad \textstyle\prod_{i=0}^{T} P(\theta_1^i\ \theta_2^i\ \dot\theta_1^i\ \dot\theta_2^i \mid X^i\ Y^i\ \dot X^i\ \dot Y^i\ \pi_2) \\
&\quad P(\ddot\theta_1^{0:T}\ \ddot\theta_2^{0:T} \mid \theta_1^{0:T}\ \theta_2^{0:T}\ \dot\theta_1^{0:T}\ \dot\theta_2^{0:T}\ \pi_3) \\
&\quad P(X^{0:T}\ Y^{0:T}\ \dot X^{0:T}\ \dot Y^{0:T} \mid X_{vp}\ Y_{vp}\ \dot X_{vp}\ \dot Y_{vp}\ I_p\ \pi_4)
\end{aligned}
$$

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

2.1 π_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, Y 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 note X_{vp} and Y_{vp} the 2D position of these via-points.

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

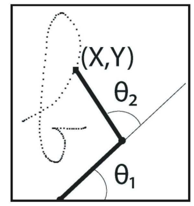

Figure 3. A two-joint manipulator, the articulatory and endpoint position variables.

the movement changes direction, and from a sensory perspective, as this means the trajectory is at a local extremum.

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

The π_1 model is defined by the joint probability distribution:

$$
\begin{aligned}
& P(X_{vp}\ Y_{vp}\ \dot X_{vp}\ \dot Y_{vp}\ I\ I_p\ Letter\ W \mid \pi_1) \qquad (2) \\
&= P(X_{vp} \mid I\ Letter\ W\ \pi_1)\ P(Y_{vp} \mid I\ Letter\ W\ \pi_1) \\
&\quad P(\dot X_{vp} \mid I\ Letter\ W\ \pi_1)\ P(\dot Y_{vp} \mid I\ Letter\ W\ \pi_1) \\
&\quad P(Letter \mid \pi_1)\ P(I \mid I_p\ \pi_1)\ P(I_p \mid\ \pi_1)\ P(W \mid \pi_1)
\end{aligned}
$$

The I and I_p variables are used as indexes in via-point sequences. The term $P(I \mid I_p\ \pi_1)$ is used to model insertions or deletions between the indexes perceived in a given trajectory I_p and the indexes in the prototypical model I. The variable W represents the person who writes the letter.

2.2 π_2: kinematic model

The kinematic model π_2 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): 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.

We define the joint probability distribution π_2:

$$
\begin{aligned}
& P(\theta_1\ \theta_2\ \dot\theta_1\ \dot\theta_2\ X\ Y\ \dot X\ \dot Y \mid \pi_2) \qquad (3) \\
&= P(\theta_1 \mid \pi_2)\ P(\theta_2 \mid \pi_2)\ P(\dot\theta_1 \mid \pi_2)\ P(\dot\theta_2 \mid \pi_2) \\
&\quad P(X \mid \theta_1\ \theta_2\ \pi_2)\ P(\dot X \mid \dot\theta_1\ \dot\theta_2\ \theta_1\ \theta_2\ \pi_2) \\
&\quad P(Y \mid \theta_1\ \theta_2\ \pi_2)\ P(\dot Y \mid \dot\theta_1\ \dot\theta_2\ \theta_1\ \theta_2\ \pi_2)
\end{aligned}
$$

The terms of the decomposition (3) describe the direct kinematic transform, which translates articulatory coordinates to endpoint cartesian coordinates. We define these terms by Dirac probability distributions. We obtain the inverse kinematic transform, which translates the endpoint cartesian coordinates to articulatory angles by inverting the direct kinematic model, using Bayesian inference.

2.3 π_3: dynamic model

The dynamic model π_3 concerns general trajectory formation for the simulated effector. It is expressed in the articulatory reference frame, and is defined by the following joint probability distribution:

$$
\begin{aligned}
& P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T}\ \theta_1^{0:T}\ \theta_2^{0:T}\ \dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T} \mid \pi_3) \\
= & \prod_{t=1}^{T-1} \left(\begin{array}{l} P(\ddot{\theta}_1^t \mid \dot{\theta}_1^t\ \dot{\theta}_1^{t-1}\ \pi_3)\ P(\ddot{\theta}_2^t \mid \dot{\theta}_2^t\ \dot{\theta}_2^{t-1}\ \pi_3) \\ P(\dot{\theta}_1^t \mid \theta_1^t\ \theta_1^{t-1}\ \pi_3)\ P(\dot{\theta}_2^t \mid \theta_2^t\ \theta_2^{t-1}\ \pi_3) \\ P(\theta_1^t\ \theta_2^t \mid \theta_1^T\ \theta_2^T\ \dot{\theta}_1^T\ \dot{\theta}_2^T\ \theta_1^0\ \theta_2^0\ \dot{\theta}_1^0\ \dot{\theta}_2^0\ \pi_3) \end{array} \right) \\
& P(\ddot{\theta}_1^0\ \ddot{\theta}_2^0\ \theta_1^0\ \theta_2^0\ \dot{\theta}_1^0\ \dot{\theta}_2^0 \mid \pi_3) P(\ddot{\theta}_1^T\ \ddot{\theta}_2^T\ \theta_1^T\ \theta_2^T\ \dot{\theta}_1^T\ \dot{\theta}_2^T \mid \pi_3)
\end{aligned} \tag{4}
$$

Inside the product over time, the first four terms model the computation of successive derivatives using finite differences, *e.g.* what are the probability distributions over velocities given the positions at time t and $t-1$, etc. The final term inside the product describes the generation of intermediary points, in the computation of trajectories between an initial position $\theta_1^0\ \theta_2^0\ \dot{\theta}_1^0\ \dot{\theta}_2^0$ and a given position to attain $\theta_1^T\ \theta_2^T\ \dot{\theta}_1^T\ \dot{\theta}_2^T$. 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 [11], where the arm first applies a maximum force, followed by a maximum negative force.

Outside the product, two terms remain, which are priors over the initial and final articulatory positions, velocities and accelerations.

2.4 π_4: vision model

We assume a simple vision model π_4, that concerns the extraction of via-points from trajectories, using their geometric properties. It is defined by the following joint probability distribution:

$$
\begin{aligned}
& P(X^{0:T}\ Y^{0:T}\ \dot{X}^{0:T}\ \dot{Y}^{0:T}\ X_{vp}\ Y_{vp}\ \dot{X}_{vp}\ \dot{Y}_{vp}\ I_p \mid \pi_4) \\
= & P(X_{vp} \mid X^{0:T}\ \dot{X}^{0:T}\ I_p\ \pi_4)\ P(\dot{X}_{vp} \mid X^{0:T}\ \dot{X}^{0:T}\ I_p\ \pi_4) \\
& P(Y_{vp} \mid Y^{0:T}\ \dot{Y}^{0:T}\ I_p\ \pi_4)\ P(\dot{Y}_{vp} \mid Y^{0:T}\ \dot{Y}^{0:T}\ I_p\ \pi_4) \\
& P(X^{0:T}\ Y^{0:T}\ \dot{X}^{0:T}\ \dot{Y}^{0:T} \mid \pi_4)\ P(I_p \mid \pi_4)
\end{aligned} \tag{5}
$$

The four first terms describe 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. The I_p variable is the index of this newfound via-point.

3 Using the probabilistic model

We have shown how π_1, π_2, π_3 and π_4, the four components of our global model π, are defined. Therefore the joint probability distribution of (1) is specified and the model is fully defined. Thanks to Bayesian inference, it can therefore be used to automatically solve cognitive tasks. We define a cognitive task by a probabilistic term to be computed, which we call a question.

3.1 Reading letters

Given a trajectory $(X^{0:T}, Y^{0:T}, \dot{X}^{0:T}, \dot{Y}^{0:T})$, what is the letter? We can recognize isolated handwritten characters if we solve the following question using Bayesian inference:

$$
\begin{aligned}
& P(Letter \mid X^{0:T}\ Y^{0:T}\ \dot{X}^{0:T}\ \dot{Y}^{0:T}\ \pi) \\
\propto & \sum_{\substack{I_p, X_{vp}, Y_{vp}, \\ \dot{X}_{vp}, \dot{Y}_{vp}, W}} P\left(\begin{array}{c} X_{vp}\ Y_{vp}\ \dot{X}_{vp}\ \dot{Y}_{vp}\ I_p\ Letter \\ X^{0:T}\ Y^{0:T}\ \dot{X}^{0:T}\ \dot{Y}^{0:T}\ W \end{array} \middle| \pi \right) \\
\propto & \sum_{\substack{I_p, X_{vp}, Y_{vp}, \\ \dot{X}_{vp}, \dot{Y}_{vp}, W}} \left(\begin{array}{l} P(X_{vp}\ Y_{vp}\ \dot{X}_{vp}\ \dot{Y}_{vp}\ I_p\ Letter\ W \mid \pi_1) \\ P\left(\begin{array}{c} X^{0:T}\ Y^{0:T} \\ \dot{X}^{0:T}\ \dot{Y}^{0:T} \end{array} \middle| \begin{array}{c} X_{vp}\ Y_{vp}\ I_p \\ \dot{X}_{vp}\ \dot{Y}_{vp}\ \pi_4 \end{array} \right) \end{array} \right)
\end{aligned} \tag{6}
$$

This question only involves terms from the representation of letters model π_1 and the vision model π_4. It can be approximated using a two-step algorithm: the vision model π_4 is first used to draw intermediary values for the positions and velocities of via-points, which are then used by π_1 to infer the probability distribution over letters.

This assumes a fully defined π_1 model: some of its terms have to be learned beforehand. More precisely, $P(X_{vp} \mid Letter\ I\ W\ \pi_1)$ and $P(Y_{vp} \mid Letter\ I\ W\ \pi_1)$ are Gaussian distributions, one for each triplet $\langle Letter, I, W \rangle$, and which are defined by their means μ and variances σ which are experimentally identified. They were learned in a supervised manner: for each triplet $\langle Letter, I, W \rangle$, the mean and the variance of the position 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.

Adding the size information in the representation of letters could surely increase the recognition rate, but there is another perspective which we would like to discuss instead. Indeed, with this question, we would like to reproduce some experimental findings, like the use of velocity cues in handwriting recognition [12]. In this experiment, subjects were shown to be able to predict the identity of the forthcoming letter: they were shown l's, and were asked to predict whether they would be followed by an e or by another l. It was shown that subjects, using the velocity information during the downstroke of the "l", had a high prediction rate. Our model could be simulated in this task, and we could quantify the role of velocity information on the recognition process.

3.2 Recognizing the writer

Who is the writer of this trajectory? Given a trajectory, our model recognizes the writer if we compute:

$$
\begin{aligned}
& P(W \mid X^{0:T}\ Y^{0:T}\ \dot{X}^{0:T}\ \dot{Y}^{0:T}\ \pi) \\
\propto & \sum_{\substack{I_p, X_{vp}, Y_{vp}, \\ \dot{X}_{vp}, \dot{Y}_{vp}, Letter}} P\left(\begin{array}{c} X_{vp}\ Y_{vp}\ \dot{X}_{vp}\ \dot{Y}_{vp}\ I_p\ Letter \\ X^{0:T}\ Y^{0:T}\ \dot{X}^{0:T}\ \dot{Y}^{0:T}\ W \end{array} \middle| \pi \right)
\end{aligned} \tag{7}
$$

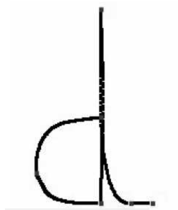

Figure 4. An example of handwriting production.

The inference is similar to (6): the question only involves terms from the model π_1 and the model π_4. However, in (6), the summation is over the variable W because we don't know the writer and we don't want to know it. In this case and for the same reasons, the summation is over the variable $Letter$. The aim of this question is to recognize the writer but not the letter.

The learning phase is done with 3 writers: the data set is composed of 3*5*26 letters. The model was then tested with 3*5*26 new letters: the rate of writer's recognition is equal to 70,9%. Misclassification arise due to the similitude of some letters for two writers.

3.3 Writing letters

Our model allows to solve the writing task, by computing: $P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid Letter\ \pi)$. What are the accelerations to apply to the arm to write a letter? We apply Bayesian inference to answer the question:

$$
\begin{aligned}
& P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid Letter\ \pi) \qquad (8)\\
& \propto \sum_{A} \left(\begin{array}{l} \prod_{i=0}^{T} P(\theta_1^i\ \theta_2^i\ \dot{\theta}_1^i\ \dot{\theta}_2^i \mid X^i\ Y^i\ \dot{X}^i\ \dot{Y}^i\ \pi_2) \\ P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid \theta_1^{0:T}\ \theta_2^{0:T}\ \dot{\theta}_1^{0:T}\ \dot{\theta}_2^{0:T}\ \pi_3) \\ P(X_{vp}\ Y_{vp}\ \dot{X}_{vp}\ \dot{Y}_{vp}\ I_p\ Letter\ W \mid \pi_1) \end{array} \right)
\end{aligned}
$$

with

$$
A = \left\{ \begin{array}{l} \theta_1^{0:T}, \theta_2^{0:T}, \dot{\theta}_1^{0:T}, \dot{\theta}_2^{0:T}, I_p, W, X^{0:T}, Y^{0:T}, \\ \dot{X}^{0:T}, \dot{Y}^{0:T}, X_{vp}, Y_{vp}, \dot{X}_{vp}, \dot{Y}_{vp} \end{array} \right\}
$$

This question involves a large summation over the set of all unknown variables, A. However, it can be approximated using a three-step algorithm: the representation of letters model π_1 is first used to draw the positions and velocities of via-points. Then, model π_2 translates these endpoint cartesian coordinates of via-points to articulatory coordinates, which are then used by π_3 to determine the trajectory between via-points. Obviously, the model π_4 is not involved in this question. Figure 4 shows a letter obtained in response to the question $P(\ddot{\theta}_1^{0:T}\ \ddot{\theta}_2^{0:T} \mid [Letter = d]\ \pi)$.

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, and the writer recognition.

Each of the sub-models leads to perspectives. For instance, the vision model might be complemented by a haptic input, which was shown to facilitate kids' learning of letters. Would our model reproduce this effect?

We could also add hierarchical layers above our letter representation, which would encode knowledge about sequences of letters, in order to obtain word recognition. This would allow to study the top-down and bottom-up mathematical propagation of information implied by our model.

In our model, we have not explicitly included constraints to force the produced trajectories to minimize some given smoothness measure. However, our "bang-bang" interpolation algorithm minimizes distances in the joint space of positions and velocities, under a constraint on the acceleration profile. How distinct are the trajectories generated by our model and by minimization procedures (jerk, torque, variance)? If they are not distinct, this raises interesting theoretical discussions.

A final perspective concerns the motor equivalence effect. In our model, "writing with the foot" can be done by replacing the motor models π_2 and π_3, by those of another effector, without changing π_1. How does this impact trajectory formation, and its quality? In other words, what part of the model would encode left-handedness?

References

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

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

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

[4] 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.

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

[6] P. Viviani and T. Flash. Minimum-jerk, two-thirds power law, and isochrony: Converging approaches to movement planning. *J. of Exp. Psych.: Human Perception and Performance*, 21(1):32–53, 1995.

[7] S. Edelman and T. Flash. A model of handwriting. *Biological Cybernetics*, 57:25–36, 1987.

[8] H.-L. Teulings. Invariant handwriting features useful in cursive-script recognition. *Fundamentals in Handwriting Recognition*, pages 179–199, 1994.

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

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

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

[12] S. Kandel, J.-P. Orliaguet, and L.-J. Boë. Detecting anticipatory events in handwriting movements. *Perception*, 29:953–964, 2000.