Optimising the Unexpected: Computational Design Approach in Expressive Gestural Interaction

Baptiste Caramiaux
Department of Computing
Goldsmiths College
University of London
b.caramiaux@gold.ac.uk

Abstract
In our work on computational design of expressive gestural interaction, we experienced various challenges for advanced optimisation methods. Here we want to highlight two of these challenges based on the design and the use of a Bayesian model called *Gesture Variation Follower*, with the aim to discuss such challenges with a broader community of designers and HCI practitioners during the workshop.

Author Keywords
Interaction Design, Gesture Expressivity, Bayesian Modelling, Optimisation, User Model

ACM Classification Keywords
H.5.2 [User Interfaces]: Interaction Styles; I.2.6 [Learning]: Parameter Learning.

Introduction
As interactive systems are now spreading outside of workspaces towards our everyday life, new elements of human behaviour such as expressivity must be embraced in technology for Human-Computer Interaction (HCI) [2]. Part of this new objective in HCI is to build technologies that are "closer", or more "natural", to human, leveraging on the use of body movements and gestures in order to enhance expressivity in interaction. Designing expressive

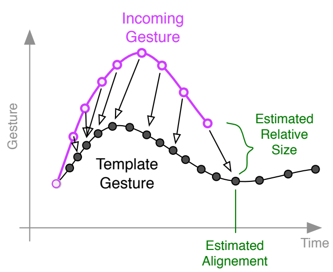

Figure 1: GVF principle: temporal alignment and tracking

gestural interaction has been the cornerstone of performing arts with digital media. For example, music technologists have leveraged on motion sensing technology to explore new types of interaction between the body and digital medias or to extend the existing ones in traditional instruments. One important challenge currently faced by artists, practitioners, researchers, designers in interaction design for the Arts is the use of complex data (provided by modern interfaces) as an expressive channel in interaction. To meet this challenge, a promising approach relies on the use of computational methods that automatically treat the data, in particular machine learning methods [5, 7].

Machine learning is a body of statistical analysis methods that achieve tasks by learning from examples. Machine learning methods are particularly useful in contexts where an application is too complex to be described by analytical formulations or manual brute force design, and when an application is dependent to the environment in which it is deployed [8].

Here we present our approach in computational design based on machine learning of gesture expressive components and present a model called *Gesture Variation Follower*. Based on the elements of modelling, we then propose to highlight challenges in terms of optimisation and HCI that we aim to discuss during the workshops.

Computational Design Approach

Our approach relies on models of user's gesture expressivity that can be implemented in interactive systems. Gesture expressivity is the notion of *how* a gesture is performed. Variation in gesture performance can exist across different users of an interactive system, or within a single user in multiple iterations recreating the same gesture. In our work, we proposed the term gesture expressivity as meaningful variation in the execution of a gesture (more details can be found in [3]). Based on this definition, our aim is twofold: to be able to recognise gestures while they are executed (**realtime classification**); and to be able to take into account the significant variations in the execution of the user's gestures by estimating them in realtime (**variations tracking**).

We identified three characteristics in the design of computational models: *temporal*, *probabilistic* with *layered representation*. Firstly, the temporal characteristic stems from the dynamic nature of gesture variations. They can occur at one specific instant (local variation) or over a longer period of time (global variation). Secondly, the captured data always involve noise. The noise can be due to the sensing system but also due to the user performing a gesture poorly. Finally, motion representation and motion variations evolve in multidimensional spaces of different dimensions. It is critical here to design models that can relate these two spaces in a simple and efficient way.

Gesture Variation Follower (GVF)

Based on the modelling strategy presented above, we conceived a model called *Gesture Variation Follower* (GVF) [4]. The model relies on two steps: learning and performing. In the learning phase, the user provides the model with examples of gestures to be recognised (*templates*). Only one example per gesture is needed. In the performing phase, the user performs a gesture and for each incoming sample the model aligns it onto the templates, computes an alignment distance and estimates the variations between the incoming gesture and the likeliest template (see Figure 1).

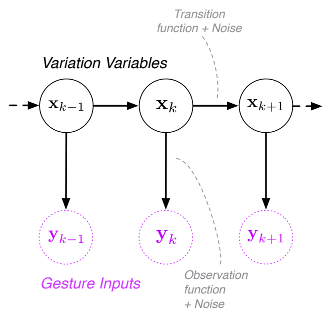

Figure 2: Graphical view of the model used in GVF

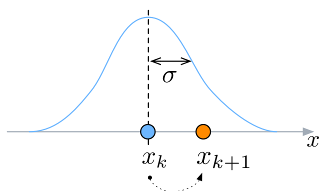

Figure 3: Transition function between state at time k and $k+1$

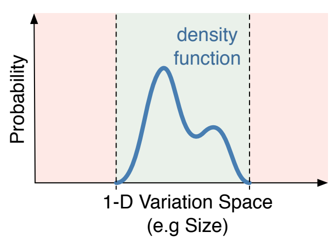

Figure 4: Exploration of variations by means of probability density function

The model is based on a general dynamical, discrete time, state-space system depicted in Figure 2. $\mathbf{x}_k$ is a vector representing variation variables that are estimated (at time k): $\mathbf{x}_k(1)$ is the alignment index; $\mathbf{x}_k(2)$ is the relative speed; $\mathbf{x}_k(3)$ is the scaling coefficient; and $\mathbf{x}_k(4)$ is the angle of rotation. $\mathbf{y}_k$ is the vector of gesture inputs (e.g. 2-d position x,y in the case of tactile surface). The model relies on a transition function from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$, set in the model as the identity plus a gaussian noise (cf. Figure 3), and an observation function which is the distance between a prediction value and the actual input. In GVF, this function is a Student's T-distribution.

The goal is to infer $\mathbf{x}_k$ based on $\mathbf{y}_k$ and $\mathbf{x}_{k-1}$. In GVF, we use a sampling method to infer gesture variations, namely particle filtering [1]. The idea behind sampling methods is to sample possible state values, i.e. possible variation values (according to a given probability distribution) and compute weights associated to each one of the values, leading to more probable values than others (and giving rise to a new distribution).

The model is open-source and available online[1]. An example of a musical application can also be seen online[2]. The idea is to start from a database of sounds. For each sound, the user records a gesture by performing it along with the sound playback (learning). In the performing phase, the user executes a gesture, the application recognises it and plays back the associated sound. In addition, variations in the execution of this gesture modulate the sound playback (speed, amplitude and filtering).

[1] https://github.com/bcaramiaux/ofxGVF
[2] http://youtu.be/2Vec40Tph5k

Optimising the Unexpected

In expressive interaction, the challenge is to adapt to unexpected variations. In this section, we propose to highlight challenges for optimisation in expressive interaction based on our work with GVF.

Exploration of Possible Gesture Variations

The capacity of the model for exploration is governed by the probability density function estimated on the space of variations. If the density function has zero-value for specific values in the variation space (or over a region), it means that these variations are "not possible" and consequently can not be considered as inputs. On the contrary, high values of the density function involves probable variations that can be taken into account (see Figure 4).

In expressive interaction, the challenge is to be able to estimate user's gesture variations without bounding the space of the possible variations in order to allow for unexpected behaviour. In GVF, the density function is incrementally estimated with a particle filter, i.e by sampling variation values (particles) over this space and by assigning a weight to each particle. The incremental estimation updates the density function at each new incoming gesture sample and consequently moves the density function towards the likeliest estimation. An advantage of the method resides in the dynamic nature of the estimation: the method adapts to new input variations. A drawback of the method resides in concentrating the search locally around the likeliest variations. While this approach has been shown to provide an accurate estimate [6], it does not allow unexpected behaviour since it discards unlikely regions of the space.

Challenge: to propose an optimisation process that allows for accurate estimation of the gesture variations and allows for the unexpected ones.

Trade-off between Latency vs. Precision
The capacity of the model to adapt to dynamic variations in the execution of a gesture is governed by its transition function (plus noise) between state estimations at t and $t+1$ that sets the dynamics of the variations. The function updates the values of the current variation estimates to the next one. The noise (often additive and gaussian) describes a neighbourhood around the updated values.

In GVF, the dynamics is essentially governed by the gaussian noise as depicted in Figure 3. In other words, the prediction of the next estimates will be done by sampling around the current values of the estimates. If the gaussian variance σ^2 is high it will then be possible to estimate abrupt changes in values (the samples will span a wider region). In this case, the model will be able to estimate rapid changes in gesture variations. However, the drawback is a loss in precision of the estimates.

Challenge: to propose an optimisation process that allows for fast convergence and accurate precision.

Contribution to the Workshop

Our goal for the workshop is to engage with researchers working on optimisation and HCI. In particular we are interested in discussing the following questions: how do the highlighted challenges overlap with challenges in other disciplines? How can expressive interaction inform on interesting computational problems for the next generations of HCI? How can these outcomes be linked with user's model of perception?

References

[1] Arulampalam, M., Maskell, S., Gordon, N., and Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing 50*, 2 (2002), 174–188.

[2] Bø dker, S. When Second Wave HCI meets Third Wave Challenges. In *Proceedings of the 4th Nordic conference on Human-computer interaction* (2006), 1–8.

[3] Caramiaux, B., Donnarumma, M., and Tanaka, A. Understanding Gesture Expressivity through Muscle Sensing. *ACM Transactions on Computer-Human Interaction (TOCHI) 21*, 6 (2015).

[4] Caramiaux, B., Montecchio, N., Tanaka, A., and Bevilacqua, F. Adaptive Gesture Recognition with Variation Estimation for Interactive Systems. *ACM Transactions on Interactive Intelligent Systems (TiiS) 4*, 4 (2014).

[5] Caramiaux, B., and Tanaka, A. Machine Learning of Musical Gestures. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (2013), 513–518.

[6] Doucet, A., Godsill, S., and Andrieu, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and computing 10*, 3 (2000), 197–208.

[7] Fiebrink, R. *Real-time human interaction with supervised learning algorithms for music composition and performance*. PhD thesis, Princeton University, 2011.

[8] Mitchell, T. M. The discipline of machine learning. Tech. rep., Carnegie Mellon University, School of Computer Science, Machine Learning Department, 2006.