Machine Learning in Expressive Gestural Interaction

Baptiste Caramiaux
IRCAM, Paris
McGill University, Montreal
baptiste.caramiaux@ircam.fr

Abstract
In our work on computational design of expressive gestural interaction, we experienced various challenges for designing accurate methods for the task at end, and their understandability and usability from the user perspective. In this proposal we present the computational approach developed in our research relying on volitional, personal, variations of gesture execution. In prior work, gesture variations have often been considered as undesirable. We present the modelling strategy undertaken to tackle this challenge and present two examples of models developed. By this proposal, we aim to discuss the encountered challenges for machine learning and human-computer interaction.

Author Keywords
Motion, Interaction, Computational Model, Machine Learning, Expression

Context
As interactive systems are 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 gestural in-

teraction 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 [6]. Music interaction can serve as an epitome of expressive gestural interaction.

Expressivity in Music Interaction

In a general understanding, expressivity involves the idea of potential variation instantiated by the consistent constitutive structure. In computer science for instance, expressivity (or expressive power) has been used in programming language theory and refers to a measure of the range of ideas expressible in a given programming language [4].

In HCI related fields, examining and designing medium for allowing expressivity is part of the core research in Music Technology, and more precisely in the NIME community, where NIME stands for New Interfaces for Musical Expression. Expression in interactive music is understood to be musical expression, connecting it to the art of all musical performance. In instrumental performance, for example, of classical music, musical expression is related directly to variation as re-interpretation of an existing piece. One pianist may interpret an established repertoire composition differently than another pianist: we can think of this as inter-user variation [8]. Or, a single performer may interpret the same composition differently in different performances: this may depend on their emotional or psychological state at the moment of a concert, the feedback the performer gets back from the audience, or through changes of context such as the size of the performance venue.

In this way, musical performance serves as a useful example of understanding expressive gestural interaction not just as an intuitive and emotional, but as volitional, contextual input to interactive systems that may facilitate human-human communication.

Challenge

Facilitating expressive gestural performance in interactive systems remains challenging. A first important challenge currently faced by artists, researchers, designers, practitioners in interaction design is the use of complex motion data (provided by modern interfaces) as an expressive channel in interaction. A second important challenge is to build systems able to understand unexpected variations of the gestural inputs. To meet these challenges, a promising approach relies on the use of adaptive computational methods that are able to automatically treat the data and being able to be adapted by the users. We argue that a promising approach relies in Human-Centred Machine Learning.

Machine learning has recently gained interest outside of its disciplines of origin (statistics and computer science), and has been shown promising for Human-Computer Interaction and Creative practices [5], especially if considering the human in the learning loop. In general-purpose machine learning, gesture variations have often been considered as undesirable and handled by (co-)variances in probabilistic learning algorithms.

In our approach, (volitional) variations in motion data are viewed as expressive vectors for interaction and, therefore, taken into account in the design strategies of our computation models.

Modelling Strategy

Our approach relies on models of user's gesture expressivity that can be implemented in interactive systems. The models are meant to extract a set of characteristics of the

physical movement execution, based on data captured from the performance. Even if we primarily mentioned the use of such models for interaction, they can serve two purposes: performance analysis and/or control parameters in interactive music systems.

Various challenges arise from analysing such performance data: the motion characteristics might not be independent of each other; the correspondence between these characteristics and the signal measured with motion capture might be non-trivial or corrupted by noise; and these characteristics may often be time-varying. Therefore, our modelling strategy relies on a probabilistic framework able to handle variability within the data through the use of time- and co-dependent random variables [7]. In particular, we propose to use a Bayesian framework, which allows for handling both variability in observed data and dependences between time-varying motion characteristics.

We present two approaches that we developed during our past investigations[1]. The first approach imposes the variation space but allows for continuous and unbounded exploration within this space. The second approach allows the user to build the variation space but bounds the exploration.

Tracking Variations
Based on the modelling strategy presented above, we conceived a model called Gesture Variation Follower (GVF) [3]. A gesture is modelled as a temporal trajectory of its characteristics and a set of variations of these characteristics along time. The set of variations is pre-defined and are typically the speed, acceleration, size, and orientation. The method uses a tracking formulation of gesture recognition under variations in order to estimate in real-time the gesture executed and its variations. The incremental tracking is using a sequential sampling method called particle filtering [1].

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. In other words, the model outputs gesture and variation estimations sample-wise.

Learning Variations
A second model aimed at learning the intended personal gesture variations provided by the performer. The goal is to ask the performer to execute the same gesture with different variations and then a model of these variations is learned. At test time, the system is able to predict the variation (or combination of variations) applied to the performed gesture.

A first version of the model uses Gaussian Mixture Models (GMM). A GMM is learned in a supervised manner where each component is a variation example provided by the user. The model has been evaluated through the case study of music conducting [9]. We found that participants with differing levels of expertise can control gesture variations of articulation and that the audio-visual feedback induces a gain in skills.

Contribution to the Workshop

Our goal for the workshop is to engage with researchers working on machine learning and HCI. In particular we are interested in discussing: how do the highlighted research question overlap with challenges in other disciplines (for

[1] Note that the implementations of the reported models can be found online at https://github.com/bcaramiaux

instance outside of creative applications)? How can expressive interaction inform on interesting computational problems for the next generations of HCI? How a human-centred approach of machine learning can help in designing expressive interaction? How allowing variations in motion performance would facilitate exploration and creativity in gesture-based interaction? How can these outcomes be linked with user's model of perception?

Acknowledgements

This research has received funding from the EU under H2020-MSCA-IF-2014 program (grant agreement no. 659232) and the European Research Council MetaGesture Music project (ERC grant agreement no. FP7-28377).

References

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

[2] Susanne Bødker. 2006. When second wave HCI meets third wave challenges. In *Proceedings of the 4th Nordic conference on Human-computer interaction: changing roles*. ACM, 1–8.

[3] Baptiste Caramiaux, Nicola Montecchio, Atau Tanaka, and Frédéric Bevilacqua. 2014. Adaptive gesture recognition with variation estimation for interactive systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 4, 4 (2014), 18.

[4] Matthias Felleisen. 1991. On the expressive power of programming languages. *Science of computer programming* 17, 1 (1991), 35–75.

[5] Rebecca Fiebrink and Baptiste Caramiaux. 2016. The Machine Learning Algorithm as Creative Musical Tool. In *Oxford Handbook on Algorithmic Music*. Oxford University Press.

[6] Eduardo Reck Miranda and Marcelo M Wanderley. 2006. *New digital musical instruments: control and interaction beyond the keyboard*. Vol. 21. AR Editions, Inc.

[7] Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*. MIT press.

[8] Caroline Palmer. 1997. Music performance. *Annual review of psychology* 48, 1 (1997), 115–138.

[9] Alvaro Sarasua, Baptiste Caramiaux, and Atau Tanaka. 2016. Machine Learning of Personal Gesture Variation in Music Conducting. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*.