Strachan, S. and Murray-Smith, R. and Oakley, I. and Angesleva, J. (2004) Dynamic primitives for gestural interaction. *Mobile Human-Computer Interaction – MobileHCI 2004: 6th International Symposium, Glasgow, UK, September 13 - 16, 2004.* 3160:pp. 325-330.

http://eprints.gla.ac.uk/2950/

Dynamic Primitives for Gestural Interaction

Steven Strachan,[1,2] Roderick Murray-Smith,[1,3] Ian Oakley,[2] Jussi Ängeslevä[2]

[1] Hamilton Institute, NUI Maynooth, Maynooth, Ireland
`steven.strachan@may.ie`
[2] Media Lab Europe, Palpable Machines Group, Dublin, Ireland
`{ian, jussi}@medialabeurope.org`
[3] Dept. Computing Science, University Of Glasgow, Glasgow, Scotland
`rod@dcs.gla.ac.uk`

Abstract. We describe the implementation of an interaction technique which allows users to store and retrieve information and computational functionality on different parts of their body. We present a dynamic systems approach to gestural interaction using Dynamic Movement Primitives, which model a gesture as a second order dynamic system followed by a learned nonlinear transformation. We demonstrate that it is possible to learn models, even from single examples, which can simulate and classify the gestures needed for the *Body Space* project, running on a PocketPC with a 3-degree of freedom linear accelerometer.

1 Introduction

Mobile telephones, Personal Digital Assistants and handheld computers are currently one of the fastest growth areas of computing and this growth is extending into fully wearable systems. Existing devices have limited input and output capabilities, making them cumbersome and hard to use when mobile. Consequently, a current requirement in this field is the development of new interaction techniques specifically designed for mobile scenarios. One important aspect of interaction with a mobile or wearable device is that it has the potential to be continuous, with the user in constant, tightly coupled interaction with the system. In these scenarios, interaction need no longer consist of an exchange of discrete messages, but can form a rich and continuous dialogue.

The Body Mnemonics project[1] develops a new concept in interaction design. Essentially, it explores the idea of allowing users to store and retrieve information and computational functionality on different parts of their bodies. In this design, information can be stored and subsequently accessed by moving a handheld device to different locations around the body. This work addresses three problem areas in mobile computing: the high levels of attention required using the devices, the impersonal nature of their interfaces, and the socially exclusive modes of interaction they support.

The work described in this paper represents first steps to providing the technology to support the gestural interaction required by the body mnemonics concept. It is concerned with developing algorithms to infer the location of a

handheld device. To provide a system that requires no additional equipment (such as worn tags or markers) to facilitate the identification of different locations, it relies on inertial sensing. Inertial sensing is a relatively new paradigm for interacting with mobile computers. Furthermore, it is a good example of continuous input; the device gathers information about user behaviour whenever it is being held or carried.

A number of researchers, such as Hinkley *et al.* [2] and Rekimoto [3], have demonstrated that inertial sensors can provide alternatives to the physical and on-screen buttons in handheld devices. They have described systems whereby shaking and tilting the device triggers different commands. However, these interfaces still possess a strong graphical component and little work has been conducted on 'screen-free' gestural interfaces. Pirhonen *et al.* [4] demonstrated a mobile mp3 player where gestures were sufficient to enable users to control the player without looking at the screen. We wish to develop the idea of screen free interaction to provide increased usability when 'on the move'.

2   Initial Explorations

Our initial investigations were conducted using an iPAQ5550 equipped with a 3-axis Xsens P^3C linear accelerometer attached to the serial port. We are concentrating on short trajectories originating and terminating at specific body locations. Several locations were considered as the source of each gesture. These were the left or right hip, where a device may naturally be held when not in use and the centre of the chest, where a device is often held to enable optimal viewing of its screen. To avoid issues of handedness, we chose to model our gestures as all originated from the centre of the chest.

Four body areas were chosen as gesture end points - left shoulder, right shoulder, back pocket and back of head. For the purposes of this exploration, all gestures were performed from the centre of the chest using the left hand whilst standing still.

The 'brute-force' approach of integrating the inertial measurements into positional trajectories and referring these to a spatial map of the body is not a strong option. A combination of uncertainty as to the precise initial position of the device and integration drift led to a substantial error margin. Figure 1 displays the trajectories inferred from acceleration measurements, for movements to the four different parts of the body, with 10 examples for each class of gesture, and makes clear the resulting inaccuracy at the end-points.

3   Dynamic Movement Primitives

The focus of this project was to choose a recognition algorithm that was flexible enough to model the required trajectories, but also constrained enough that it could be trained with minimal effort, using a small number of example gestures by a novice user.
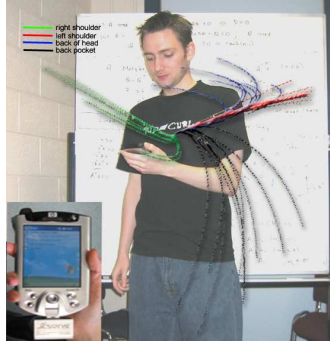
Fig. 1. Example of the drift encountered when acceleration traces are integrated into positions. Significant integration drift is observed, leading to end-point uncertainty.

The *Dynamic Movement Primitives* (DMP) algorithm proposed by Schaal *et al.*, is *"a formulation of movement primitives with autonomous non-linear differential equations whose time evolution creates smooth kinematic control policies"* [5, 6]. The idea was developed for imitation-based learning in robotics, and is a natural candidate for application to gesture recognition in mobile devices. It allows us to model each gesture trajectory as the unfolding of a dynamic system, and is better able to account for the normal variability of such gestures. Importantly, the primitives approach models from origin to *goal* as opposed to the traditional point-to-point gestures used in other systems. This, along with the compact and very well-suited model structure enables us to train a system with very few examples, with a minimal amount of user training and also provides us with the opportunity to add richer feedback mechanisms to the interaction during the gesture.

DMP's are linearly parameterised enabling a natural application to supervised learning from demonstration. Gesture recognition is made possible by the temporal, scale and translational invariance of the differential equations with respect to the model parameters.

A Dynamic Movement Primitive consists of two sets of differential equations, namely a canonical system, $\tau \dot{x} = h(x)$ and a transformation system, $\tau \dot{y} = g(y, f(x))$. A point attractive system is instantiated by the second order dynamics

$$\tau \dot{z} = \alpha_z(\beta_z(g - y) - z), \ \tau \dot{y} = z + f, \tag{1}$$

where g is a known goal state (the left shoulder, for example), α_z and β_z are time constants, τ is a temporal scaling factor, y and $\dot{y}$ are the desired position and velocity of the movement and f is a linear function approximator. In the case of a non-linear discrete movement or gesture the linear function is converted to a non-linear deforming function

$$f(x, v, g) = \frac{\sum_{i=1}^{N} \psi_i w_i v}{\sum_{i=1}^{N} \psi_i}, \ \text{where} \quad \psi_i = e^{-h_i\left(\frac{x}{g} - c_i\right)^2} \tag{2}$$
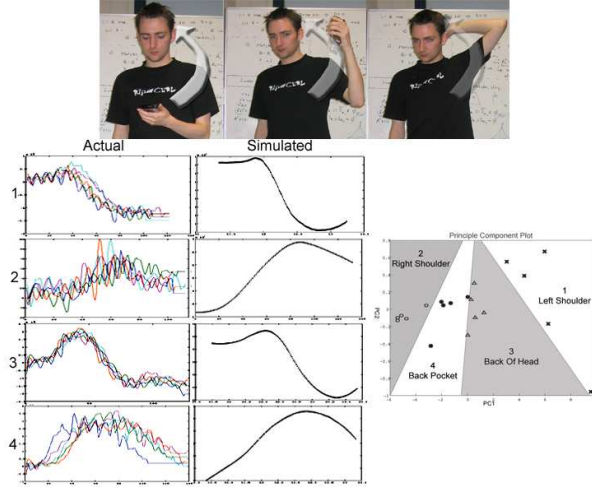
Fig. 2. Five realisations of the four gestures on the x-coordinate are shown along with an example simulated gesture from the DMP model. A principal component plot shows the separability of the model parameters. Similar results can be demonstrated for the y and z coordinates also.

These equations allow us to represent characteristic non-linear behaviour that defines the gesture, while maintaining the simplicity of the canonical 2nd order system driving it from start to goal. The transformation system for these discrete gestures is

$$\tau\dot{z} = \alpha_z(\beta_z(r - y) - z) + f, \ \tau\dot{y} = z, \ \tau\dot{r} = \alpha_g(g - r) \tag{3}$$

where $\dot{z}$, z and y represent the desired acceleration, velocity and position respectively.

The approach to learning and predicting the dynamic movement primitive is to provide a step change in reference and pass this through the non-linear deforming function. Values for the f's can be calculated along with sets of x's and v's from the canonical system and this is then passed through a Locally Weighted Projection Regression (LWPR) algorithm [7] that learns the attractor landscape and allows us to make predictions of the function f given values for x and v.

4 Results, Future Work and Conclusions

Our implementation of the Schaal DMP algorithm, running on a pocket PC with inertial sensing, provides the basis for an efficient, robust and rapidly trainable gesture recognition system for the four basic gestures we tested.

Figure 2 shows examples of acceleration time-series corresponding to the x-coordinate acceleration trace for each class of gesture, along with the simulated

curve from the learned model in the second column. The good match between the measured and simulated curves provides encouraging evidence of its suitability for gesture recognition, especially as each simulated gesture was generated using a model trained on only one example of the five shown, and in only five iterations of the LWPR algorithm. The separability of the model parameters for classification purposes is visible in the plot of the first two principal components for each of the four classes of gesture.

The additional benefit of this dynamic approach is that it provides the designer with the opportunity to incorporate rich, continuous feedback mechanisms into the interaction with the user. We can now deliver continuous audio or tactile feedback relating to the user's motion, proximity to goals, or gesture trajectories [8]. We believe this kind of tightly coupled control loop will support a user's learning processes and convey a greater sense of being in control of the system. For this we will be using the MESH hardware platform [9], which features a 3-axis accelerometer, 3-axis gyroscope, 2-axis magnetometer and an integrated vibro-tactile transducer with a large (54 dB) dynamic range. The richer sensor input will broaden the scope of interaction possibilities, and the system features the dynamic vibrotactile output required to display the probabilistic feedback from our DMP models.

References

1. Ängeslevä, J., Oakley, I., Hughes, S., O'Modhrain, S.: Body Mnemonics. In: Mobile HCI Conference 2003, Udine, Italy. (2003)
2. Hinckley, K., Sinclaire, M., Horvitz, E.: Sensing techniques for mobile interaction. In: ACM Symposium on User Interface Software and Technology, CHI Letters 2 (2. (2000) 91–100
3. Rekimoto, J.: Tilting operations for small screen interfaces. In: ACM Symposium on User Interface Software and Technology. (1996) 167–168
4. Pirhonen, A., Brewster, S., Holguin, C.: Gestural and Audio Metaphors as a Means of Control for Mobile Devices. In: Proceedings of ACM CHI2002 (Minneapolis, MN), ACM Press, Addison-Wesley (2002) 291–298
5. Schaal, S., Peters, J., Nakanishi, J., Ijspeert, A.: Learning Movement Primitives. In: International Symposium on Robotics Research (ISRR2003). (2004)
6. Ijspeert, A., Nakanishi, J., Schaal, S.: Learning attractor landscapes for learning motor primitives. In S. Becker, S.T., Obermayer, K., eds.: Advances in Neural Information Processing Systems 15. MIT Press, Cambridge, MA (2003) 1523–1530
7. Vijaykumar, S., Schaal, S.: Locally Weighted Projection Regression: An $O(n)$ Algorithm for Incremental Real Time Learning in High Dimensional Space. In: 17th Inter. Conf. on Machine Learning (ICML2000), Stanford, California. (2004)
8. Williamson, J., Murray-Smith, R.: Granular synthesis for display of time-varying probability densities. In Hunt, A., Hermann, T., eds.: International Workshop on Interactive Sonification (Human Interaction with Auditory Displays). (2004)
9. Oakley, I., Ängeslevä, J., Hughes, S., O'Modhrain, S.: Tilt and Feel: Scrolling with Vibrotactile Display. In: Proceedings of Eurohaptics 2004, Munich Germany. (2004)