Strachan, S. and Murray-Smith, R. and O'Modhrain, S. (2007) BodySpace: inferring body pose for natural control of a music player. In, *CHI '07 Extended Abstracts on Human Factors in Computing Systems, April 28-May 3 2007*, pages pp. 2001-2006, San Hose, California, USA.

http://eprints.gla.ac.uk/3514/

BodySpace: inferring body pose for natural control of a music player

Steven Strachan[1]
[1]Hamilton Institute
NUI Maynooth
Maynooth, Co. Kildare, Ireland
Steven.Strachan@nuim.ie

Roderick Murray-Smith[1,2]
[2]Dept. of Computing Science
University Of Glasgow
Glasgow, Scotland
rod@dcs.gla.ac.uk

Sile O'Modhrain[3]
[3]Sonic Arts Research Centre
Queens University
Belfast, N.Ireland
sile@media.mit.edu

CHI 2007, April 28 – May 3, 2007, San Jose, USA
ACM 1-xxxxxxxxxxxxxxxxxx.

Abstract

We describe the BodySpace system, which uses inertial sensing and pattern recognition to allow the gestural control of a music player by placing the device at different parts of the body. We demonstrate a new approach to the segmentation and recognition of gestures for this kind of application and show how simulated physical model-based techniques can shape gestural interaction.

Keywords

Gesture Recognition, pattern recognition, accelerometer, music player, interaction design

ACM Classification Keywords

H.5.2. User Interfaces: Interaction Styles; H.5.2. User Interfaces: Haptic I/O; H.5.2. User Interfaces: Input devices and strategies;

Introduction

The Body Mnemonics project [1] developed a new concept in interaction design, harnessing the ancient 'method of loci' technique. Essentially, it explores the idea of allowing users to store and retrieve information and computational functionality by moving a handheld device to different locations around the body. Moving the device to the back pocket, for example, may open a

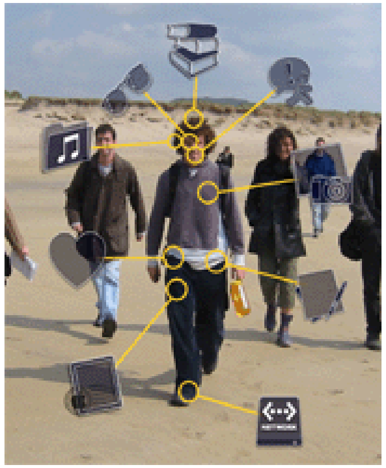

Figure1: Examples of other ways we may use the bodyspace system

user's personal finances on the mobile device. From the machine's sensor perspective this is simply a time-series in acceleration but the user thinks of the back pocket as the location of their personal finances, and that this can be accessed by moving their phone there.

Previous work on this concept focussed mainly on the basic ideas and requirements for the project without a working, mobile implementation. Strachan *et al.* [12] built a dynamic systems implementation of a gesture recogniser. In this paper we describe the first implementation of a completely handheld and fully functioning 'BodySpace' system, which uses inertial sensing to recognise when it is placed at different areas of a user's body in order to control a music player, essentially utilising the human body as the mnemonic device. A user may place the device at their hip in order to control the volume of their current song or at their ear, in order to switch tracks.

Our system differs from other gesture controlled systems in that we are not required to explicitly design a lexicon of gestures. The range of gestures we use is constrained by the limits (static and kinematic) of the human body in that the arm can only move to a finite number of locations around the body, providing us with an obvious, perfectly natural and easily generated set of gestures. Another difference is that we do not use any buttons at all in our interface, making the interaction more fluid and natural than a gestural system that requires an explicit button press at the beginning of each gesture (e.g. the Samsung SCH 310, the only gesturally controlled phone on the market, uses a gesture button which has to be activated while generating gestures). Additionally, we also use a model-based approach to our interaction design enabling us to easily alter the dynamics of interaction and multimodal feedback by varying the parameters of our model.

Gesture Controlled Applications

Inertial sensing has proved to be a viable technique for sensing movement for gestural interaction with mobile devices. Rekimoto *et al* [9] describe their GestureWrist system, which consists of a wristband that recognises hand and forearm movements and uses these movements to communicate with a computer. Ubi-Finger [13] is another system which uses acceleration and touch sensors to detect a fixed set of hand gestures and Kela *et al* describe the use of a matchbox sized sensor pack, SoapBox [3], which they use to control the functionality of different appliances in their design studio. They describe a study designed to compare the usefulness of the gesture modality compared to other modalities for control such as RFID objects or PDA and stylus, finding that gestures are a natural modality for certain tasks. This reflects the conclusions of Pirhonen *et al.* [7] who investigated the use of gesture and non-speech based audio as a way to improve the interface on a mobile music player. The key advantage of this gestural approach is that it enables eyes-free interaction with a music player, which is advantageous, especially when the user is `on the move'.

Hardware

The equipment used consists of an HP iPAQ 5550 running windowsCE equipped with a MESH [6] inertial navigation system (INS) backpack consisting of 3 Analog Devices ±2g dual-axis accelerometers, 3 Analog Devices ±300deg/s single chip gyroscopes, 3 Honeywell devices magnetometers, a Trimble LassenSq GPS and a

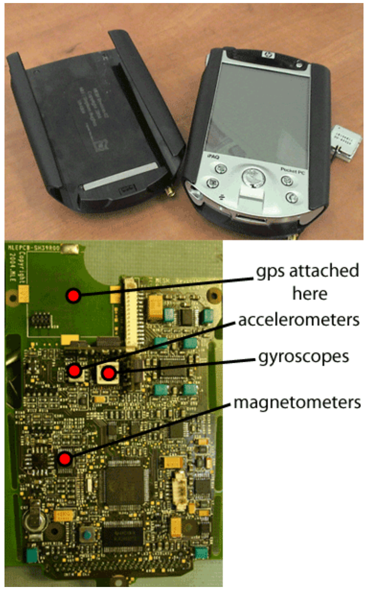

Figure3: MESH inertial sensing backpack

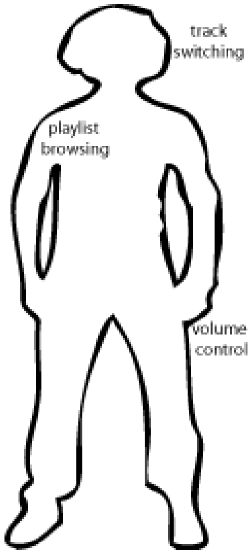

Figure4: Body locations with functionality

vibrotactile transducer providing us with an unprecedented level of information about the current inertial state of our mobile device.

Gesture Recognition And Segmentation

One of the major challenges for any continuously sensing system is how do we know when to interpret sensor readings as control signals, rather than background activity? How do we detect what is meaningful? And how do we detect user intention? [5] The detection of user intention has been investigated by Powers [8] who illustrated examples of intentional behaviour which could be empirically detected using continuous control models. Williamson and Murray-Smith [15] described an interface built on this principle. Utilising methods from perceptual control theory and dynamic systems they present a method for performing selection tasks based on the continuous control of multiple, competing agents who attempt to determine the user's intentions from their control behaviour enabling users to select an object without the explicit use of a pointer. One important issue which needs to be addressed is the mismatch between what a user perceives the system to be doing and what the system is actually doing, referred to as an *isomorphism error* [14]. In our case the user perceives the system to be checking the position of the device with respect to the body but in reality what the device is doing is monitoring angles and pattern matching accelerometer data so it is important that we attempt to reduce the effect of this isomorphism error on our system.

Our Approach

One of the aims of this project is to avoid the use of explicit button presses to segment or separate one gesture from another, as this tends to interrupt the natural fluidity of user interaction with the system. Our approach to the recognition of when the device is placed at different body parts is a two-stage process. The first stage involves identifying if the device *may be* at a certain part of the body, which we refer to as the *Segmentation Stage* and the second stage involves the classification of the accelerometer data immediately prior to the notification from the segmentation stage, the *Recognition Stage*. It is important then that we adequately represent the state of our system.

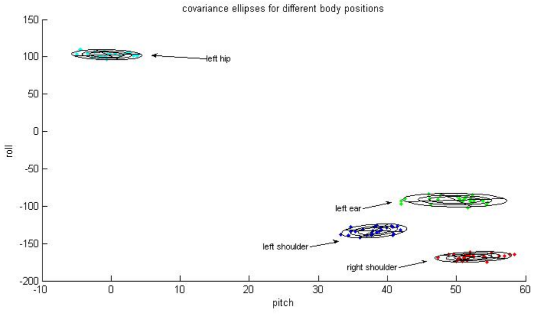

Figure 2: four covariance ellipses corresponding to each part of the body where pitch and roll data was measured.

The state vector could contain any information relevant to the action to be inferred. In our case we may use information from our accelerometers or from our gyroscopes, which allow us to monitor the general movement of the device be that rotation, larger translational movements or tremor from our muscles [11] to model the state of the system at the goal point of each gesture. We chose, in this configuration, to represent the state of our device in a simple way using its orientation, pitch and roll pairs, at the end point of a gesture observed as the device was placed at different parts of the body. Pitch and roll pairs, from

Figure6: Illustration of the main functionality of the BodySpace system.

accelerometer measurements, can indicate which body locations are, as shown in figure 4, compatible. This then allows us to use the second stage of recognition to gather extra evidence for the inferred body location by comparing how the device moved to that position, based on accelerometer data for the last second of motion. A simple Multi-Layer Perceptron [2] is used to classify one of four body positions. The use of a Multi-Layer Perceptron at this stage shows the generality of the approach and was perfectly adequate for this task since its compact final form of a handful of parameters, and low processing cost makes it very suitable for low memory mobile devices. Training of this system involves repeated gestures to the four different parts of the body with three gestures per location required to achieve adequate training in this set-up.

Modelling

In this work we also incorporated dynamic model-based approaches to interaction. By basing our interaction on a simulation of a physical model we enable a more active exploration of the potential range of interaction with the device. It also allows us to alter the 'look and feel' of the interaction very easily by simply altering the parameters of the model and gives us great scope for designing multi-modal interaction, where the vibration and audio feedback can be generated in real-time as the user performs the gesture.

When the device is classified to be at a certain body location, the system switches to the correct mode and model associated with that part of the body. So for example, when we wish to switch tracks, the device is first moved to the left ear where recognition occurs. A mode switch then means that when the device is tilted back or forward at the ear, in order to switch tracks, as in figure 6, a simulation of a 'ball in a bowl' metaphor represents the state of the interaction. We can imagine a ball placed in a bowl or concavity as shown in figure 5 where each bowl or concavity represents a different track. The simulation approach allows us to add formative feedback such as the real-time synthesis used by Rath & Rocchesso in [9] which gives the user a sense of the devices sensitivity to his action.

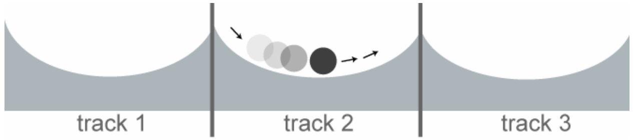

Figure 5: Combination of bowls, which the user must navigate the ball into in order to switch tracks

With a row of bowls representing a list of tracks, it is possible simulate the task of transferring a ball from one bowl to the next by providing an external force from the movement of the device. In this case the external force comes from a flick of the device. Increased velocity and momentum of the flick would allow users to reach the peak, and effectively fall into the next track. We may model the surface friction and the effort required to overcome the peak of the bowl with some simple physics. Each bowl is represented by a simple parabola, with a certain height, y, used to calculate angle of slope: $\theta = \tan^{-1}(x/y)$ and the force: $F = mg\sin\theta$, minus surface friction [4]. This interaction is also augmented with vibrotactile feedback allowing the user to feel when the track switch has occurred, where the level of feedback presented is associated with a parameter of the physical model. A similar mechanism is used to control the volume of a track, which is located, in this set-up, at the left hip. So

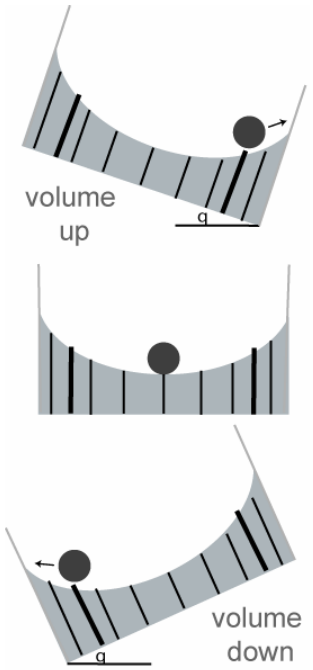

Figure 7: When the ball rolls into the left side of the bowl the volume decreases. When it is rolled into the right side of the bowl the volume increases

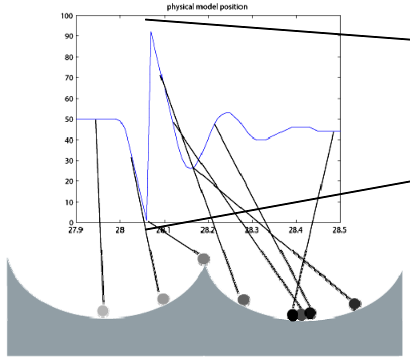

Figure 10: Example of how the position output data maps to the position in the simulated bowl

when the device is placed at the left hip the mode switches to a volume control mode. This mode is represented by only one bowl as shown in figure 7 so that when the device is held level there is no change in the volume but when the device is tilted the ball rolls to one end of the bowl over a number of lines, each representing a vibrational pulse. At the end of the bowl the ball is stopped and a larger vibrational pulse is felt by the user. This approach is useful, not only because it offers an intuitive way of providing feedback, but also in cases where there may be increased general movement, such as noise from walking movements or from being inside a vehicle. This context could be detected by the system, which could then alter the dynamics of the model. For example the bowl could become larger when the user is walking or the movement of the ball on the surface of the bowl could become more viscous making false-positive track switches or volume changes much less likely to occur in that context.

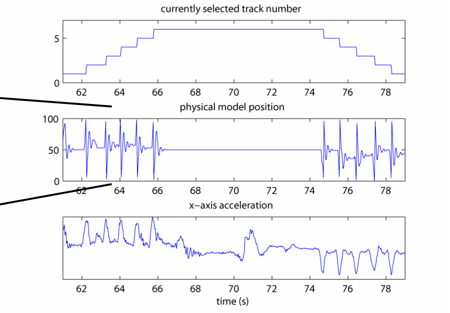

Figure 9: Example of the data observed for a track switching. Here the track is moved forward five times then back again five times.

Example

Figures 8 and 9 show examples of how the accelerometer data interacts with our simulated physical model. Figure 9 shows how accelerometer data provides the energy to the model, which switches the current track by causing the ball to roll into the next bowl as illustrated in figure 10. Figure 8 shows that as the device is tilted, in volume control mode, the ball in the physical model rolls to the edge of the bowl. If enough energy is imparted to the ball, it will pass a threshold, causing the volume to increase or decrease depending on the direction in which the ball rolls.

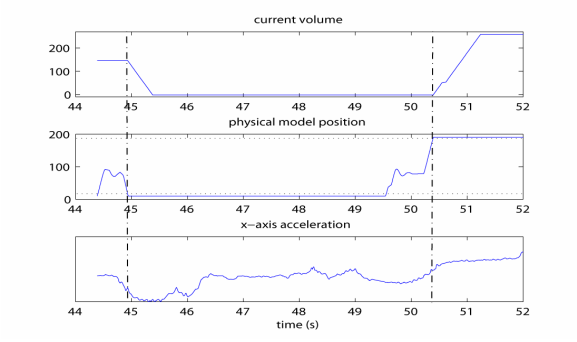

Figure 8: Example of the data observed for a volume control. Here the volume is first decreased then increased again.

Discussion and Conclusions

We described a handheld system that utilises inertial sensing and basic pattern recognition to allow the gestural control of a music player by simply placing the device at different parts of the body, rather than having to press buttons, or dial wheels. The system also does not require the user to wear instrumented clothing – it could be used on the beach. We have demonstrated a new approach to the segmentation and recognition of gestures for this kind of application and that a model-based approach to this kind of interaction can be both intuitive and enables the easy provision and adjustment of feedback.

The control of a music player is just one potential application for this system but it may also be used for other tasks such as the retrieval or storing of files or the activation of different functionalities at different parts of the body as in figure 1. You may also wish to call your girl/boyfriend just by placing the device at your heart or answer the phone by placing the device at your ear.

The social acceptability of a system such as this is very important and must be considered at the design stage. It is generally accepted that input devices should be as discreet and natural as possible, which has been a significant problem with previous gesture-based systems, which were considered too obtrusive or too obvious. It is essential that any future evaluation of this system looks from two different angles: 1) usability and 2) acceptance of the interaction technique on the part of the user and of observers. Although our gestural examples in this case are very extroverted for illustrative purposes, it is simple to alter the system to function with more subtle gestures, and the control could be via a sensor separate from the device controlled (e.g. a sensor in a watch or headset for controlling a phone, via Bluetooth).

Acknowledgements

We are grateful for support from: HEA project Body Space, SFI grant 00/PI.1/C067, IRCSET BRG SC/2003/271, Continuous Gestural Interaction with Mobile devices, EPSRC project Audioclouds, IST Programme of the European Commission, under PASCAL Network of Excellence, IST 2002-506778.

References

[1] Ängeslevä, J. Oakley, I. Hughes, S. and O'Modhrain, S. Body Mnemonics. Mobile HCI Conference 2003, Udine, Italy. ACM, 15–22.

[2] Bishop, C. M. Neural Networks for Pattern Recognition. Oxford University Press, 1995.

[3] Kela, J., Korpipää, P., Mäntyjärvi, J., Kallio, S., Savino, G., Jozzo, L. and Marca, D. Accelerometer-based gesture control for a design environment. Personal and Ubiquitous Computing. 10, 5 (2006), 285–299.

[4] Kelly, J. Video of tilt-menu for mobile phone emulator. http://www.dcs.gla.ac.uk/rod/Videos/JamesMenu2.avi, 2004.

[5] Hinckley, K., Pierce, J., Horvitz, E., and Sinclair, M. Foreground and background interaction with sensor-enhanced mobile devices. ACM Trans. on Computer-Human Interact. 12, 1 (2005), 31–52.

[6] Oakley I. Ängeslevä J. Hughes, S. and O'Modhrain, S. Tilt and feel: Scrolling with vibrotactile display. EuroHaptics 2004.

[7] Pirhonen, A. Brewster, S. and Holguin, C. Gestural and Audio Metaphors as a Means of Control for Mobile Devices. : Proc ACM CHI2002 (Minneapolis, MN). ACM Press, Addison-Wesley (2002), 291–298.

[8] Powers, W. T., Behavior: The Control of Perception. Aldine, Hawthorne, NY, 1973.

[9] Matthias Rath and Davide Rocchesso. Continuous sonic feedback from a rollling ball. IEEE Multimedia, 12(3), 2005.

[10] Rekimoto, J. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices, Fifth International Symposium on Wearable Computers, 2001.

[11] Steven Strachan and Roderick Murray-Smith. Muscle Tremor as an Input Mechanism. UIST, 2004.

[12] Strachan S. Murray-Smith R. Oakley I. and Ängeslevä J. Dynamic Primitives for Gestural Interaction. Mobile HCI 04. Springer (2004).

[13] K Tsukada and M Yasamura. Ubi-finger: Gesture input device for mobile use. APCHI '02.

[14] Williamson, J. Continuous Uncertain Interaction. PhD thesis, Dept Computing Science, University Of Glasgow, 2006.

[15] Williamson, J. and Murray-Smith, R. Pointing without a pointer. CHI '04: CHI '04 extended abstracts on Human factors in computing systems. ACM Press 1407–1410.