SUPPORTING GESTURAL INPUT FOR USERS ON THE MOVE

R Headon and G Coulouris [rph25,gfc22]@cam.ac.uk

University of Cambridge, England

ABSTRACT

A wearable system is described that enables users to perform simple command selection and input operations. The system is suitable for use while moving and in almost any posture. The implementation uses RFID and is low in cost and power requirements. There is provision for changing between application contexts, making the system suitable for use in mobile and ubiquitous computing environments where the mobility of the user requires frequent changes of focus. The paper includes a preliminary evaluation of the system's usability.

1 INTRODUCTION

Ubiquitous and wearable computing systems are increasingly interactive. While the sensing of location and other contextual information enables applications to operate more autonomously than in desktop systems [1], there remain many situations in which explicit commands and selections must be provided by users. Techniques are required to support interaction with portable, wearable and fixed ubiquitous computing devices.

Potential applications range from the control of a slide presentation during a lecture to the use of public services such as travel guides, timetables and ticket purchase, the operation of an in-car or head-mounted mobile phone, digital camera or personal stereo as well as many others in work, leisure, travel and other contexts. Users may be seated, standing, walking or in other postures, in situations where the use of a keyboard, device for cursor control or even a touch-sensitive panel for input is impractical. This paper describes a prototype for a wearable gesture input device designed for use in such situations and its integration into ubiquitous computing applications where the presence of a display cannot be assumed.

We have established the following design requirements:

- A method for selecting between options that will be determined by the context of use.
- Suitable for use when standing, walking, or even riding a bicycle.
- Feedback to the user during selection (to ensure accuracy) and after (for closure) without visual distraction.
- Low cost and ubiquitous.

These requirements lead to the following design decisions:

- Input selection by arm gesture between a small number of options (achievable number to be determined by experiment).
- Conveniently wearable with low power requirements.
- Auditory feedback.
- A software framework is required to support context (application) switching and translate gestures into input streams for applications while providing user feedback and maintaining user awareness of the current context.

These design decisions led to an implementation using RFID (radio frequency identification) and a software framework designed to support interaction. Gestures are performed by positioning the right arm relative to the body. A reader antenna is worn on the user's wrist and passive RFID tags are worn on the body. Context defines the application to which gestures are communicated. Explicit context switching is usually achieved by the user selecting a tag attached to an object or poster associated with the required application. The antenna is connected to an RFID reader which polls any tags that are within reading range. The RFID reader is mounted on an evaluation board that supplies a serial interface and is placed in a shirt pocket and powered by an external battery. An iPaq is connected to the serial port to initiate poll operations, receive tag observations and provide audio feedback and instructions to the user when required. The iPaq uses a wireless network to transmit tag readings to a server that runs the software framework. The framework translates the tag values to appropriate commands or context changes and transmits the commands to the currently active application (figure 1).

Section 2 describes support for two levels of gestures and context switching. A practical application and scenarios are presented in section 3. Section 4 describes the implementation of the wearable sensor. As an early evaluation, section 5 presents an experiment conducted to determine the accuracy of tag selection while walking. Section 6 presents related work. Section 7 concludes.

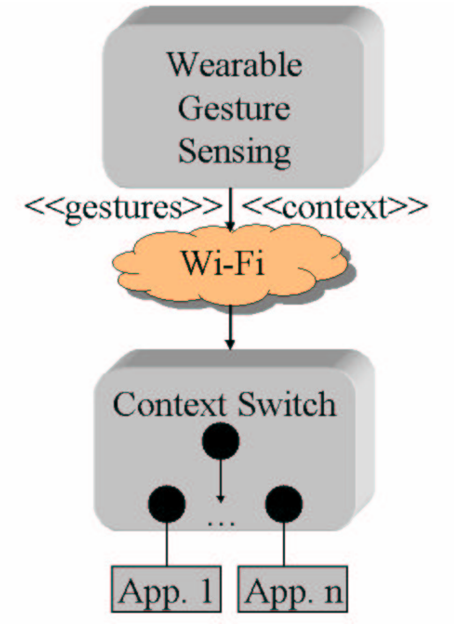


Figure 1 – Gesture streams are sent from the wearable over a wireless network to the context switch that decides which application should receive the gesture input stream.

2 CONTEXT SWITCHING AND GESTURE INPUT

Gestures are recognised by the location of the wrist (reader antenna) relative to a region of the body (tag). Sequences of these static gestures produce dynamic gestures. Each gesture has a specific meaning in the application *context* and represents the input to an active application. The user's *explicitly* selected context defines the application.

Figure 2 – Example of a gesture.

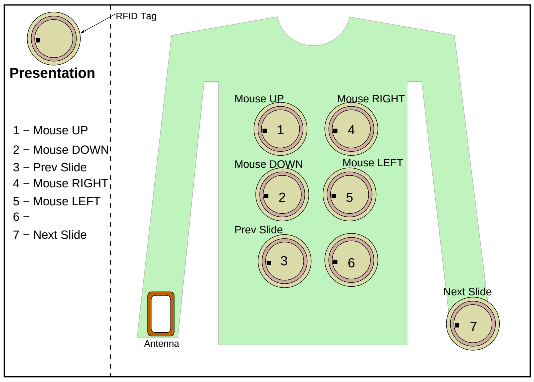
Static and Dynamic Gestures

The body is segmented into regions which generate a gesture event when the wrist enters this region. Each region is defined by the location of a passive RFID tag and the range of the wearable RFID reader (antenna). There is a small null space between regions to disambiguate them. Tags can be placed anywhere to define new regions, however for gesture input it is convenient to have these regions defined primarily on the front of the torso within easy reach.
A gesture is indicated in figure 2 which also shows the tag placement in the middle of the shirt (large white squares). On the torso there is a 3×2 array of tags, with an additional tag placed on the left wrist attached to the shirt cuff or watch. The tags are evenly spaced either side of the centre occupying an area within easy reach and with small displacements to transit between tags. However additional tags are also supported and their meaningful placement, such as one near the ear, may represent a command to select audio feedback. To gesture, the user simply brings the reader antenna, attached to the wrist, within reading range of a tag attached to the shirt. The antenna-tag range is approximately 75mm in the current implementation. A tone corresponding to that tag is then played as user feedback. The frequency of the tone corresponds to the spatial arrangement; higher tags (i.e. closer to the head), have a higher frequency. When a dynamic gesture is entered, the audio feedback consists of the tones representing the individual tags being played back in sequence.

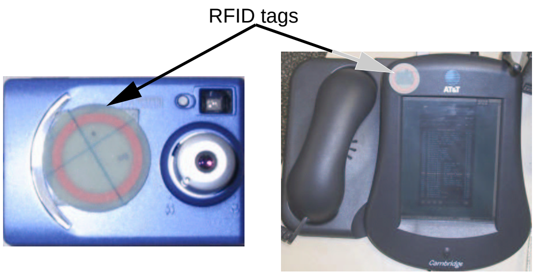
Context Switching

It is necessary to supply a mechanism to support smooth change of context so as not to confine the application of gesture input, e.g. how to terminate control of the personal stereo and assume control of the digital camera. This is analogous to starting a application on your desktop and then directing input (keyboard and mouse) focus to that application.
For mobile users, context can sometimes be inferred from their location, but in general, the user needs the option to explicitly select an application context. Context can be specified in a similar way to gestures, and an RFID tag represents a context selector.
Context selector tags are placed on posters and objects, or may even be specified as a gesture. This does not imply you need to stay at the location of the poster or object to affect interaction, rather this launches an application to which gestures are directed. The user remains mobile, maintaining the command set until another explicit change in context occurs. A **poster** (figure 3(a)) consists of a sheet of paper to which a context selector tag is attached and a schematic list of gestures together with a corresponding command are printed. In some circum-

stances a **physical object** (figure 3(b)) is a better representation of context.



(a) Poster



(b) Objects

Figure 3 – Context selectors.

3 APPLICATIONS AND SCENARIOS

The aim is to provide a practical and usable gesture input system for stationary and non-stationary users. One example where mobility is advantageous is during a presentation where the speaker may like to control the slide sequence and pointer, while maintaining the freedom to wander the room. This section describes the implementation of this application. However there is great potential for an eyes-free, non-contact, "no fingers", portable gesture input device supporting the actively mobile user outside the office domain into almost every industry involving computer interaction or input. This section presents some of these application scenarios.

Controlling a presentation was chosen as an early application as it is a practical application and concurrent demonstration for researchers presenting a gesture input system. The "presentation" context is selected using a poster close to the display; either a projector screen, desktop monitor or shared public display. On selecting this context the application teleports the user's desktop to this screen using VNC [2]. The user can then load the presentation as usual, or if a paper copy of the presentation has a tag attached, select this to display the presentation. The poster associated with the presentation context also depicts the gesture-command mapping. Gestures are used to control the sequential selection of slides, and also to control cursor motion. The presentation application interprets gestures to generate key events and mouse events to the window manager that displays the slides.

The gesture system was also used in an application to navigate through a virtual world and control a music player on a computer. A context was defined for each application with a poster to show the gesture-command mapping.

A non-contact, eyes-free, wearable input device has application scenarios that exploit the features of this input mechanism. Many people carry a small digital camera everywhere, such as the one in figure 3(b). Consider the situation of cycling past a "capture the moment" scene and wanting to take a photograph, without stopping. It is awkward and difficult to remove the camera from the pocket and then control it. If the camera was head-mounted, an RFID tag representing the camera context could initiate gestures to control the device, including taking a picture. The same approach can be applied to other personal devices that require control while moving, such as a telephone, personal stereo or PDA.

4 IMPLEMENTATION

An RFID system is a sensible choice for a wearable device due to its small size, low cost, low power and contact-less properties. An antenna is worn on the wrist, with the reader evaluation board comfortably placed in a shirt pocket with a cable connecting the two. RFID tags are placed on a shirt. The sensing of a tag gives the the location of the wrist *relative* to the body, thus being invariant to the user's location, orientation and posture (e.g. bent over). The location also exhibits high resolution, defined by the read range on the antenna/reader package. This section describes the hardware shown in figure 4.

The Shirt and Tags

RFID tags are integrated with clothing to facilitate the recognition of gestures. The tags are passive, i.e. all the energy is drawn from the reader interrogation signal. Tag size is a critical factor in determining comfort and noticability. Afixing tags to the inside of the shirt maintains its original appearance. It is desirable that the wearer does not feel a difference in the garment when wearing the tags. The tags are disc shaped with a diameter of 30mm and thickness of 0.7mm. The thinness of the tag makes them almost unnoticable to the user. An ordinary shirt, which even has metal buttons, was chosen for easy donning and exchange between users.

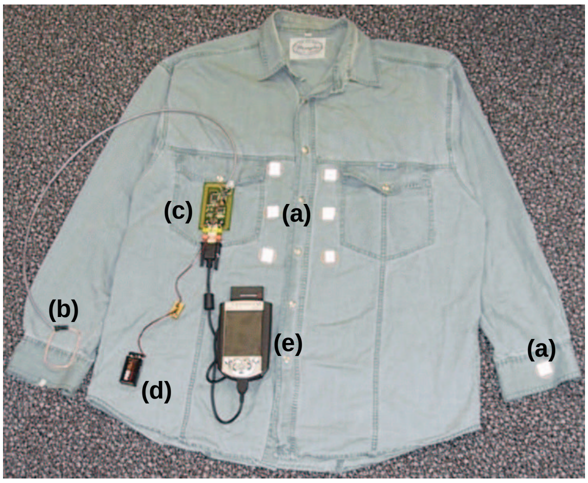


Figure 4 – Wearable gesture input system with components exposed for clarity. The large white squares identify tag locations (a), the antenna is placed on the right-hand wrist (b), the reader evaluation board on the right shirt pocket (c), battery (d) and iPaq (e) on the bottom right of the shirt.

Wearable Reader

The reader should be unobtrusive, comfortable to wear and use, and consists of three parts: reader chip, RF circuitry and antenna. This prototype uses a low-frequency (125kHz) RFID evaluation board from EM Microelectronic [3]. There is no requirement for the antenna to be adjacent to the reader chip. Initially the antenna was attached to the cuff of the shirt, however during movement the antenna would not correctly reflect the orientation of the wrist. The antenna is therefore attached to the user's wrist using a strap.

The antenna design determines the read range of the system. The tags are passively powered, hence the orientation of the tag with respect to the reader antenna also affects the read range. As the antenna is worn on the wrist it is desirable to make it small and still maintain a sufficient read range. Initial antennas were circular with a diameter of 45mm, to fit completely within the width of a wrist, an inductance in the range of 350μH to 750μH, and a low Q. However the read range was not sufficient and not robust to tag orientation. The second set of antennas were rectangular, with a dimension of 50mm $\times$ 70mm, an inductance of around 350μH and a low Q. This antenna gives a read range of approximately 75mm when the tag is parallel, and is robust to tag rotation. The antenna is an air coil, formed by turns of 0.2mm enamelled copper wire bound closely together using Spiraband, and connected to the reader using shielded cable. This antenna was deemed sufficent, if a little large, for initial trials.

Power, Control and Communications

The reader evaluation board is controlled using a serial connection over which commands and tag sightings are reported. To make this device wearable, or at least portable, an iPaq is used to to drive the reader and provide audio feedback. Many applications reside not on the wearable computer itself, rather on the servers which can respond to and control the environment. The iPaq is equipped with a wireless network card and regurgitates the tag observations on a TCP port to which interested applications connect. The reader evaluation board requires a 5V supply which comes from a 5V 500mA regulator connected to a 9V battery.

Feedback

Feedback is required to acknowledge gestures, both static and dynamic. Visual feedback is not appropriate as the device is intended for eyes-free interaction. Auditory feedback is chosen where each tag has an associated tone (frequency) played when that tag is observed by the reader. The iPaq is used to play these tones and the user typically wears an audio earpiece in one ear only.

Vision

The prototype, while certainly wearable, is a combination of components that were flexibly designed to operate for a range of applications. The vision is to integrate the power, sensing (reader and antenna), control and communication (iPaq) components into a single integrated device. This device could be the size of a wristwatch, and equally robust. The device would consist of a microcontroller, reader chip, RF module, antenna, communications module (e.g. bluetooth) and audio feedback. The tag sightings would be communicated locally to wearable devices and to fixed basestations for server-based applications.

5 EXPERIMENT

An initial investigation was undertaken to evaluate the accuracy of absolute and relative tag selection while walking. In the absolute selection experiment the user is instructed to return their right arm to the side prior to each tag selection. In the relative case the user maintains wrist proximity close to the tag from the previous selection.
The user puts on the wearable gesture system. Each user runs through some training sequences to become comfortable and familiar with the location and feedback sounds selecting the tags produce. The experiment starts with the user walking down a long corridor and they are prompted to select a specific tag through the earpiece. The sequence and times of tag selections are recorded. When the correct tag

is found they are instucted to locate another tag. There were three participants and each experiment consisted of nine trials (tag selections). The results for the two conditions are presented in table 1.

TABLE 1 – Results.

Condition	#trials	#errs	%err.	Mean time
Absolute	36	4	11.1%	1.72s
Relative	24	3	12.5%	1.17s

An error is identified when the user did not select the correct tag on the *first* attempt. The majority of selection errors and significant selection time is noted at the start of each experiment. Although the selection error is quite low, it is expected to reduce with longer trials and increased user training. The mean time to correct selection requires further investigation as the interval between attempts to read a tag was long at one second.

6 RELATED WORK

Many gesture recognition systems exist that use computer vision, inertia trackers or touch sensitive surfaces to sense and recognise gestures. Many of these implementations are not portable, nor has their application while "on the move" been explored. This section focuses on portable gesture sensing devices.

Gesture input for users on the move is challenging. Brewster [4] approaches this problem with a multimodal approach combining nod-shake detection, and finger gestures input to a touch sensitive iPaq display. Head movement is observed using an inertial tracker and the user enters gestures to the iPaq worn on a belt.

Perng [5] reports a device that observes static forces acting at the finger tips. This accelerometer data is transmitted wirelessly to a server that performs hand gesture recognition. The user wears a glove with wires running from the power, control and communication unit on the wrist to the accelerometers mounted at the finger tips. Another wireless inertial measurement unit for gesture recognition is presented by Benbasat [6]. However inertial devices are sensitive to vibration noise from the environment (e.g. on a bus) and created by human movement.

A wearable RFID tag reader, although not wireless, is reported by Schmidt [7]. The antenna is sewn into a glove which the user wears providing a maximum read range of 70mm. The purpose is to trigger computer actions in response to the user handling tagged physical objects, for example to load a webpage when the user holds a pen.

Non-gesture devices, such as keyboards and speech can be used to control wearable devices. The Twiddler [8] is a one-handed keyboard often used on wearables, though it requires training and can be difficult to use. The device is sizable, expensive and requires a dedicated hand and fingers. Speech recognition does not work well in noisy environments and is quite intrusive in the presence of others.

7 CONCLUSION

This paper presents a new gesture input device and context switching support suitable for use while on the move. Gestures are sensed as the *relative* position of the wrist to the body and recognition is invariant to the user's location, orientation and posture. The implementation exhibits good accuracy while walking and further experiments are expected to yield better results. The proprioceptive nature of the device explains the ability for eyes-free operation and accuracy is expected to improve with user training. Operation is non-contact which makes it suitable for conditions where finger usage is inappropriate, for example in a clean-room where contamination by touching objects should be minimised, in medical applications where the hands are too dirty or too clean to touch a device, or for users with reduced mobility in their fingers. The user's function is not impeded as the hands remain free to hold items, although the system is not usable when both of the user's arms are used simultaneously for balance or support. Close proximity sensing using RFID is robust to changes in environment conditions. Many other systems use gestures for input that have a social meaning, such as nodding a head or waving an arm. The gesture set presented in this paper uses close and central body gestures making its use discrete.

RFID technology is low power, low cost, small sized and simple. Gestures are recognised as a single tag sighting, or sequence of tags. Complex recognition algorithms are not required and the system does not require calibration or per user training. Further experiments are planned with longer trials and to evaluate the accuracy of dynamic gestures while on the move. However the gesture system will also be evaluated using applications to determine accuracy of tag selection and user feedback in a practical application.

ACKNOWLEDGEMENT

The authors thank Andy Hopper, Frank Hoffmann, Ian Wassell and those who endured the experiment.

REFERENCES

[1] Hopper, A., 2000, "The Royal Society Clifford Patterson Lecture, 1999 – Sentient Computing", Phil. Trans. R. Soc. Lond., 358, 2349–2358

[2] Richardson, T. et al., 1998, "Virtual Network Computing", IEEE Internet Computing, 2, 33–38

[3] "EM Microelectronic – Marin SA". http://www.emmarin.com

[4] Brewster, S. et al., 2003, "Multimodal 'Eyes-Free' Interaction Techniques for Wearable Devices", Computer Human Interaction, ACM

[5] Perng, J. K. et al., 1999, "Acceleration Sensing Glove (ASG)", International Symposium on Wearable Computers

[6] Benbasat, A. and Paradiso, J., 2001, "An Inertial Measurement Framework for Gesture Recognition and Applications", International Gesture Workshop, Springer-Verlag

[7] Schmidt, A., Gellersen, H., and Merz, C., 2000, "Enabling Implicit Human Computer Interaction A Wearable RFID-Tag Reader", International Symposium on Wearable Computers

[8] Barfield, W. and Caudell, T., 2001,"Fundamentals of wearable computers and augmented reality", Lawrence Erlbaum Associates