

Gestures and multimodal input

Dr Simeon Keates

Dr Peter Robinson

Bionotes

Dr Simeon Keates is a Research Associate in the Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB3 0DG, UK. Dr Peter Robinson is a Lecturer in the Computer Laboratory, University of Cambridge, New Museums Site, Pembroke Street, Cambridge. CB2 3QG, where he is part of the Rainbow Group working on computer graphics and interaction. He is a Fellow and Director of Studies in Computer Science at Gonville & Caius College. He is also a Fellow of the British Computer Society and a Chartered Engineer.

1. Abstract

For users with motion impairments, the standard keyboard and mouse arrangement for computer access often presents problems. Other approaches have to be adopted to overcome this.

In this paper, we will describe the development of a prototype multimodal input system based on two gestural input channels. Results from extensive user trials of this system are presented. These trials showed that the physical and cognitive loads on the user can quickly become excessive and detrimental to the interaction. Designers of multimodal input systems need to be aware of this and perform regular user trials to minimise the problem.

2. Introduction

Gestural interfaces are gaining popularity, because they are intuitive and can be tailored to a particular individual (Wolf 1986). This is of particular benefit to a user with impaired motion.

If a gestural input system is well designed, the transducer used for generating the data for the gesture recognition process can also be used to control the cursor. This effectively removes the need for ancillary equipment, such as switches, and means that users only have to deal with a single input device, thus saving the time and effort involved in using multiple devices for the interaction process.

2.1 The benefits of multimodal input

Any computer input system intended for use by people with varying physical disabilities and designed around one method of input is unlikely to be flexible enough to cope with the diverse needs and demands of the users satisfactorily. This is not to say that it might not suffice, but for extended computer usage something more flexible may be required.

There is evidence to suggest that increasing the degrees-of-freedom of input devices, such as incorporating finger flexion, can improve interaction rates (Zhai et al. 1996). Extending this principle to include more degrees-of-freedom through multiple input channel implies that this should also yield improved information transfer rates.

It is the inherent flexibility of a multimodal input system that is its main advantage. If one particular mode of input is difficult to use, the simple substitution of another, easier to use mode that is readily available within the input system should alleviate the problem. Using a number of modes can increase the vocabulary of symbols available to the user. Increased flexibility of input should mean more efficient user-computer interaction and information transfer through an increase in the available bandwidth of communication.

3. Developing A Multimodal Prototype

A prototype system was developed for investigation of the interaction processes involved in multimodal input. The system was intended to be as intuitive as possible to operate. As with any unfamiliar interaction process, there is a potential for very high cognitive loads. These cognitive overheads can negate any potential gain from reducing the physical load on the user, so the system had to be easy to learn.

When considering the available input modes, the temptation is to optimise the input system for a particular target application. This is not a satisfactory solution because it leads to designing a new input system for each application. Also, the widespread use of consistent designs for software interfaces minimises the need for optimising for particular software packages.

3.1 Selection of input modes

The input modes for the multimodal prototype were chosen to be complementary, so they could be combined without producing adverse effects on the interaction process. For the purposes of the prototype system, the input channels used were two gestural input channels, the head and the hand.

There has been an increase in interest in applications of gestural input in the rehabilitation field (Keates *et al.* 1997). Using gestures in the prototype allowed a more comprehensive overview of gesture interaction to be developed.

There were also usability advantages to using two gestural input channels, using similar formats for the input modes limits the cognitive load on the user as similar motions are required for each mode.

At least two transducers are required for multimodal input. For the prototype, a 3Space Polhemus was used for head gesture recognition, as the users had previously experienced using this input mode and device in an earlier set of trials (Keates *et al.* 1997). A standard analogue games joystick was used for the second mode, being similar in nature to the wheelchair joysticks with which many of the users were familiar.

Although the specific input modes used for the prototype were head and hand gestures, these are a representative and not a definitive set of input modes. They are not being presented as a possible universal input system for all potential users. The system used was designed to gather empirical data about the interaction process.

3.1.1 The gestures

Head gestures are governed by the nature of the neck, allowing ball and socket type swivelling, but limited translational movement.

(1) Simple directional gestures (figure 1)

These gestures consist of the movement of the head along the principal axes of rotation. There are four gestures of this type for the head: UP, DOWN, LEFT and RIGHT. An UP gesture is produced by tilting the head backwards, increasing its angle of pitch. A DOWN is produced by pitching it forwards towards the chest. LEFT involves rotating the head to the left by yawing rather than rolling. Similarly, a RIGHT is generated by a yaw in that direction. Rolling was not used because it is not sufficiently distinct from yawing for reliable recognition. Each of these gestures involved a return to the central datum position, that is where the head is naturally at rest when facing the screen, for completion.

Insert figure 1 about here

(2) Oscillatory gestures (figure 2)

There are two oscillatory gestures, YES and NO. These consist of combinations of the simple directional gestures. A NO involves several LEFT and RIGHT gestures to produce a shaking of the head. For improved recognition, it is necessary to combine at least three of these gestures. The YES is performed by a series such as UP-DOWN-UP, mimicking the natural nodding form of assent.

Insert figure 2 about here

Jester, a software application developed within the University of Cambridge, was used for the gesture recognition. The hybrid recognition algorithm used in Jester is based upon a strategy utilising both dynamic time warping techniques and heuristic rules (Perricos 1996). Dynamic time warping has been used successfully in speech recognition, and the recognition of hand gestures. It is an optimisation technique whereby the vector distance between two finite, time varying signals is calculated, after the difference due to varying signal lengths has been minimised.

Heuristic rules were developed based on knowledge about the particular modes of motion of the head, such as limited translational movement. These are used to classify the gestures, before they are recognised by the dynamic time warping algorithm. This has the advantage of reducing the number of templates that need to be searched through, thereby increasing the recognition speed.

3.2 Control strategies for combining the inputs

Three principal ways in which multimodal input can improve the interaction process were investigated:

1. offering a range of modes and allowing the user to select the most appropriate input channel, with no interaction between the modes (Single Mode);

2. repeating the same input data on several channels synchronously to increase the recognition accuracy (Both Modes Same);
3. processing the input data from a number of channels that is independently variable on each channel and can be combined to increase the available vocabulary size (Both Modes Different).

Each of the above strategies were assessed through extensive user trial sessions.

3.3 Implementing the control strategies for multi-channel independent input

When using several independent recognition systems, an implementation strategy is required to coordinate their channels of information. Two options for this are neural networks and statistically-based heuristic rules. The latter approach was adopted for the prototype to minimise the training data needed.

The simplest control strategy implemented for repeated data on multiple channels (Both Modes Same) was a heuristic based on a ‘simple majority’, with any unclear situation, such as two contradictory inputs, classified as not recognised. If one input channel produces a result of ‘not recognised’ then that channel is disregarded. Table 1 shows how such a heuristic would work for three vertical gestures, YES, UP and DOWN.

Insert table 1 about here

With several separate recognisers, this method can have a detrimental effect on the recognition rate as high quality information is indiscriminately contaminated with more spurious data and there is no compensation for this.

Subtler heuristic approaches proved more effective than the simple majority approach. The heuristics developed were based on generic input properties derived from the data obtained from the interaction process and optimised accordingly.

For example, studying the data obtained from the initial evaluation of Jester, it was clear that the premature segmentation of the oscillatory gestures, YES and NO, was one of the more frequent causes of recognition errors. The converse, simple directional gestures being mistaken for oscillatory ones, was less frequent. Therefore, one possibility for a more complex heuristic was to assume that if an oscillatory gesture was produced on one channel, and an appropriate directional one on the second channel, then the resultant gesture should be the oscillatory one. This heuristic is summarised in table 2.

Insert table 2 about here

One problem with using gestural input is unwanted or unintended bodily movement, such as sneezing or another person distracting the user's attention. There are two ways of coping with such inputs. One method is to have each input mode filter out such movement independently of each other. The other is to develop a generic control strategy that is robust enough to deal with such occurrences.

This can be achieved by filtering at a higher level using plan recognition, the prediction of intent by the computer through user interaction models (Smart et al. 1993), in which the computer expects particular inputs at certain times and weights the recognition process accordingly. An example of this prediction would be waiting for a single or double click command when activating a button.

The former strategy was adopted for the prototype, because Jester is sufficiently robust to filter out most erroneous input, having been designed for standalone operation and therefore not dependent upon other algorithms to filter out the input noise.

4. Method

The most appropriate criterion on which to test the efficacy of a new input system is the rate of information transfer between the user and the computer, as the target of any input system is to maximise this. This involves minimising the following:

- 1- user planning time;
- 2- time per input;
- 3- system response time;
- 4- correcting input errors (for example, user produces wrong response or system interprets input incorrectly).

To provide a ready measure of the success of the multimodal system, some form of benchmarking is required. The most straightforward method of achieving this is to examine the relative efficiency compared to that of the standalone recognisers. The first step in this process is the calculation of the recognition rates obtained by existing systems. The overall recognition of the multimodal input system should be no less than the lowest of these, and should aim to be at least as high as the highest.

4.1 *Multimodal user trials*

A series of user trials was established for a period of three months with the users described in table 3. A multimodal system prototype was used, recognising head and hand gestures.

Insert table 3 about here

4.2 *Aims of the user trials*

The aims of the user trials were to identify generic traits of multimodal input, to evaluate the feasibility of using gestures as at least one of the input modes, and to provide empirical data about the human-computer interaction process for motion-impaired users.

Insert figure 3 about here

4.2.1 Software

For the user trials, software was developed to provide prompts for the user to produce multimodal input. The program, the Gesture Input Test, was designed to study the production of gestures under differing cognitive and physical loading.

As can be seen from figure 4, the gesture prompts were presented verbally on screen, and simultaneously read out by the computer operator to provide an auditory prompt as well. This combination of visual and auditory cues maximised the likelihood of the user understanding the required gestures first time, effectively increasing the intensity of the stimulus that the user was being exposed to, and reducing user uncertainty.

Insert figure 4 about here

A gesture vocabulary of six directional gestures was chosen: UP, DOWN and YES, as defined for the 3 gesture test, as well as LEFT, RIGHT and NO. These gestures were chosen to minimise the cognitive load on the user. It was thought that being primarily directional in nature, the relative minimisation of abstraction of the gestures should help with the users' understanding of what was being asked for when prompted. To assist in this reduction of cognitive load on the users, a visual cue, similar to figure 5, was drawn and positioned at the bottom of the screen.

Insert figure 5 about here

There was the option for utilising different inputs modes singly or in combination. Two input modes were used: head gestures via a Polhemus and hand ones using an analogue joystick.

Tests were completed with each mode individually (Single Mode - SM) and combined with both inputs being specified as either different gestures (Both Modes Different - BMD) or identical (Both Modes Same - BMS) to reflect each of the control strategies discussed earlier. For the BM Same mode, the users were encouraged to produce both gestures simultaneously, whilst for BM Different they were given free choice of either simultaneous or sequential input. The original intention had been for this mode to be simultaneous only, but it became clear at a very early stage in the trials that this would be simply too difficult, so the test was modified. Exposure to the tests was staged in the order Single Mode, BM Same and BM Different so as to coincide with what was perceived as the increasing cognitive load. Trial sessions were conducted on a basis of one hour per user every alternate week, subject to the availability of the users.

4.3 Results from the Papworth user trials

If gestures are to be used as a potential input system for everyday computer use, then it is not only the recognition rates that are of concern, but also the misrecognition and non-recognition rates. A gesture that is not recognised requires repetition of the input, but a gesture which is misrecognised needs to be corrected before the input can be repeated. To reflect this difference in the required number of corrective inputs, a scoring system was implemented.

Insert figure 6 about here

Any gesture correctly recognised earned +1 point, a gesture which was not recognised scored 0 points, and an incorrectly recognised gesture received a -1 score. These scores were then scaled to a maximum score of 100 to eliminate dependence on the number of gestures recorded per test. This gave a complete

potential score range of -100 to +100. By definition, any score below 0 would be entirely unacceptable, because the user would spend all the time trying to correct errors, and could never produce any useful input beyond the simplest of tasks.

Figure 6 shows comparative scores generated for Single Mode input with a vocabulary of three gestures, YES, UP and DOWN, and those for a 6 gesture vocabulary.

Figure 7 shows the resulting graph from the full Gesture Input Test results for the 6 gesture vocabulary. There are three generic groups of data shown. The first of these is the score from Single Mode (SM) entry, where only one gesture is prompted for on one input channel. This was the first task experienced by the users. Both Modes Same (BMS) was the second one. Here the user was faced with two identical gesture prompts and asked to produce them simultaneously on the two input modes. Finally, Both Modes Different (BMD) shows the scores generated when the prompt consists of two different gestures. Each of these data sets is subdivided between the head and hand component scores. The data is not well suited to statistical significance analysis, but the comparative scores obtained for each input mode were consistent across all users.

Insert figure 7 about here

The first of the BM Same lines shown, 'BMS Head/Hand', are the direct recognition scores seen by the head/hand inputs with no account taken of the other mode. This gives a direct comparison of user performance under different cognitive and physical loads when compared to the Single Mode results. The second of the two lines shows an optimised combination of the head and hand data. This was derived from a set of heuristic rules that were refined as the trials progressed, based on the numerical data gathered and the observations made during the user sessions. Several strategies were examined.

The first of these was taking one or other of the input modes as the primary input and only referring to the secondary one when no result was recognised. However, this offered no chance of correction of misrecognised gestures, simply reduction in the non-recognition rate.

Taking a simple majority of the two inputs gave some scope for error removal. If the two input modes agreed, that was the gesture taken as the recognised one. If either input was classified as not recognised then the other mode result was taken. Finally, if the two modes disagreed then the final result was classified as not recognised. This is identical to the rule base described in table 1.

This yielded improved scores, and further refinement was made when it was observed that when the oscillatory gestures (YES and NO) were performed, there was a tendency for the recogniser to segment the gestures prematurely and recognise them as purely directional ones (LEFT, RIGHT, etc.). This behaviour is illustrated in figure 8(a) with the YES and NO being classified primarily as the directional components of those gestures. Figure 8(b) shows the equivalent results for Single Mode head input.

Insert figures 8(a) and 8(b) about here

This was corrected by applying a new rule that if the two modes disagreed then if either mode recorded an oscillatory gesture that would be taken as the gesture result. This was further modified by assuming that if the modes recorded LEFT and RIGHT gestures together, then the intended input was a NO. Attempts to apply this to UP and DOWN being YES did not yield any improvement in results owing to the confusion as to which way was UP on the joystick.

This merits some explanation. Despite using joysticks for driving wheelchairs, the instinctive choice of UP was actually DOWN for most users, effectively treating the joystick as an aeroplane one, rather than a games or wheelchair one. This effect was mitigated to some extent by the use of a visual cue, but was

still noticeable. The line labelled 'BMS - Opt' in figure 7 shows the resultant scores when both modes were combined using these rules.

Finally, the two single points on figure 7 'BMD - Head/Hand' represent the results from the single session of Both Modes Different entry. This was the final task that the users were introduced to. Only a single session was available for this, before the completion of the trials.

The Gesture Input Test times recorded were for the complete process shown in figure 9. Again the physical gesture time could be calculated from the number of samples and transducer sampling rate.

Insert figure 9 about here

The times recorded for the BM Different results are the total times from the appearance of the prompt to the completion of the recognition process for the respective input channel. Consequently, the apparent discrepancy for the BM Different Hand columns is because virtually all the users opted to do the head gesture first, and the hand gesture second. This means that the BMD Hand results are inclusive of the time to complete the head gestures as well and correspond to the times to complete the input on both channels. However, simply calculating the difference between the two sets of results will not yield the time taken for the hand gesture, because there is a degree of overlap in the cognitive processing required between the two gestures requested that cannot be evaluated explicitly.

Figure 9 shows that the time to complete the recognition of the gestures, is approximately 50% longer for the two oscillatory gestures than for the purely directional ones. This is a combined effect of the nature of the gestures and the template based recognition. The YES and NO gestures are compounds of three of the directional ones. Therefore, they should take approximately three times as long to physically perform. However, the Hybrid Recognition Algorithm (Perricos 1996) looks for specific features, including how many times the transducer passes through the zero or datum position. Once this happens

at least as many times as the particular user threshold, twice for all the users here corresponding to the completion of the second of the three gestures, the computer recognises it as an oscillatory gesture, and terminates the data retrieval as soon as it identifies which oscillatory gesture. Hence the physical movement time lies somewhere between 2 and 3 times that for a purely directional gesture.

4.4 Evaluation of the results

Looking first at figure 6, which shows the comparative performance between the 3 gesture vocabulary and the 6 gesture one, it can be seen that the 3 gesture vocabulary consistently outscores the 6, with the single exception of the third session. Given that the extra three gestures are similar in nature to the original three and, if anything, more reliable as inputs, then the only explanation for discrepancy is the effect of having to remember, decide between and cognitively process the six gestures.

Under BM Different input, the users eventually ceased to look at the screen prompts, choosing instead to rely solely on the aural prompt from the computer operator reading the required gestures aloud. They later reported that this was because it was easier to remember the instructions when spoken. This enabled the users to recall the entire instruction without having to continually refer to the information source and has implications for future multimodal systems if the users become so overloaded by the interaction process that they cannot produce inputs without constant double-checking of the prompt.

The scores generated by the Gesture Input Test for the 6 gesture vocabulary are shown in figure 7. The highest scores overall came from the Single Mode head input, and show a typical learning curve as predicted by the Power Law of Practice (Card et al. 1983). That Single Mode Head scored more highly than Single Mode Hand is probably because the users were already familiar with head gestures from a previous set of trials, but that effect was virtually cancelled out by the 5th session, when the users were becoming equally familiar with both.

Both the Single Mode inputs scored more highly than their direct counterparts in the Both Modes Same tests. This is likely to be as a result of two influences. One is that the physical nature of the gestures

was altered slightly by trying to produce two physical motions with the body at the same time, and these generated interference of some kind in each other (the templates were trained with each mode being used separately). This is possibly a significant source of error, although it cannot be readily compensated for without increasing the level of recogniser training required by the square of the number of gestures as the noise effect from a gesture will be depend upon which gesture it is.

The other likely cause is that the cognitive load on the user, from remembering the task, learning a new process and controlling a wider range of muscle groups simultaneously, had increased from the original Single Mode test.

By careful analysis of how the errors were arising, a heuristic rule base could be developed to combine the data from both inputs to combine scores that were much closer to the single mode ones.

The Both Modes Different trial was similar to the Single Mode ones, no more than one gesture was being produced at any moment in time, but the cognitive load was increased, because the user was responding to more complicated instructions. It can be seen from figure 7 that the BM Different scores were lower for both input modes when compared to the Single Mode score from the same session. This is the fairest comparison, because the users were significantly more familiar with generating gestures by this stage, so comparison to earlier sessions would not give an accurate reflection.

Analysing the scores, whilst giving a valuable insight into the relative reliability of the particular mode combinations, does not fully describe the data transfer rates achieved as no account is made of the relative vocabulary sizes involved. To rectify this, the bit rates of useful information transfer, can be calculated from equation 1:

$$Rate = \frac{\log_2(Vocabulary_size) * \frac{Score}{100}}{Time_taken} \quad (1)$$

For each of the Single Mode and Both Mode Same the vocabulary size was 6 gestures, whilst for Both Modes Different it was 36 gestures (6 * 6). Using the above equation to calculate the data bit rates

achieved, the Single Mode head and hand inputs achieved peak rates of 0.72 and 0.77 bits/second respectively. This compares to the 0.65 bits/second obtained for BM Same and 0.56 bits/second for BM Different.

5. Discussion

The objective of the user trials with the multimodal input prototype was to establish whether such input is feasible and if it offers any benefits over single mode systems.

That a prototype was developed and used to obtain results shows that such systems are technically feasible, although the prototype was far from an operational configuration. The processing requirements for a system based on several gestural input channels are such that considerable processing power is required, but with the advent of faster chips and multi-processor desktop computers, that no longer presents a problem.

The most immediate result of this work is the demonstration that, in principle, gestures could form a component part of a computer input system. This is particularly, but not exclusively, useful for users with motion impairments.

The actual benefit depends on how the modes are being used. For instance, if both modes are being used concurrently to reinforce each other, that is to produce the same gesture on multiple channels rather than one and take the average of the results, then this actually yields poorer results than using a single mode only. This is because the physical and cognitive load on the users becomes such that the integrity of the gestures produced is adversely affected.

Learning is possibly a major factor in the results seen. Multimodal input is a difficult process to learn and this imposes significant cognitive demands on the user, particularly in conjunction with the production of unfamiliar combinations of physical movements.

If the modes are being used to generate different gestures then this cannot be done simultaneously as the load on the users is simply too great. Therefore, given that the recognition accuracy of two gestures

produced on different channels is less than two gestures produced on the same channel, and that the times involved are comparable, then the only way of obtaining extra value from the second channel is to make the gestures mean different things according to the channel that they are produced on. For the prototype used here, this would mean that a joystick LEFT produced a different effect from a head LEFT. This method would increase the amount of information that a user could produce within a given period of time by up to a factor of the number of modes.

As figure 7 shows, it is clear that Single Mode entry, particularly for head gestures, is apparently superior to any of the multimodal hybrid systems for reliability. This is because the higher cognitive load associated with the planning and physical motion necessary for multimodal input affects adversely the quality of input generated by the user. An analysis of the bit rates generated for each of the different input strategies used during the user trials shows that the potential gains in interaction rate through increased vocabulary size are not realised for the same reasons. Indeed, the data transfer rates actually decrease.

There are two possible methods for overcoming this deficiency in performance. The first is to change how the modes are used. Instead of trying to combine the data from the different input channels, they could be treated as entirely separate, the recognition process ceasing on complete recognition of a gesture on any of the input channels. This would, in effect, turn the multimodal system into a collection of many single-mode input channels.

The alternative method to improve the performance of the multimodal system is through developing a training regimen for the users to increase familiarity with generating the required movements. Such a training scheme is contrary to the intention to make the interface intuitive and natural, however it is not without precedent. Operating chord keyboards involves pressing several keys at the same time. Fast input rates are possible, but many hours of training are required to become proficient (Cushman and Rosenberg 1991).

6. Conclusions

These results show that multimodal input using gestures is feasible. However, there are a number of drawbacks. The biggest sources of difficulties are undoubtedly the physical and cognitive loads (including the learning demands) placed on the user. Any multimodal system, therefore, needs to be very carefully designed and tailored to each individual user to ensure that at no point do these loads become excessive.

7. Acknowledgements

The authors would like to thank the staff of the Bradbury Progression Centre and the users at the Papworth Trust for giving their time so freely and generously. We would also like to thank the late Robin Jackson for providing the inspiration for this work.

8. References

- Card, S. K., Moran, T. P. and Newell, A. 1983, *The Psychology of Human-Computer Interaction*, (Lawrence Erlbaum Associates, New Jersey).
- Cushman, W. H. and Rosenberg, D. J. 1991, *Human Factors in Product Design*, (Elsevier Science Publishing), 193 - 200.
- Keates, S., Potter, R., Perricos, C. and Robinson P. 1997, Gesture Recognition - Research and Clinical Perspectives, *Proceedings of RESNA '97*, (20 - 24 June) (RESNA, Pittsburgh), 333 - 335.
- Perricos, C. 1996, Jester - A Head Gesture Recognition System for Windows 95, *Proceedings of RESNA '96*, (7 - 12 June) (RESNA, Salt Lake City), 304 - 306.
- Smart, W. D., Flensholt, J., Gomard, C., Cairns, A. Y., Ricketts, I. W. and Hollnagel E. 1993, ARCHIE - Plan Recognition in a Human-Computer Interface, *Proceedings of the 12th UK Planning/Scheduling SIG Workshop* (6 - 7 May) (Cambridge, UK).
- Wolf, C.G. 1986, Can People Use Gesture Commands?, *IBM Technical Report #86A002284*.

Zhai, S., Milgram, P. and Buxton, W. 1996, The Influence of Muscle Groups on Performance of Multiple-Degree-of-Freedom Input, *Proceedings of ACM CHI '96* (13 - 18 April) (ACM, Vancouver), 308 - 315.

Please note: Frames and pictures are centre-aligned, Word tables are left aligned.

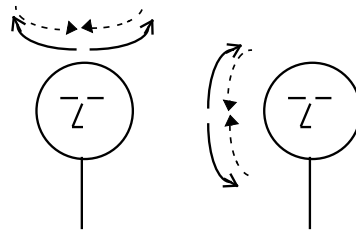


Figure 1. LEFT/RIGHT UP/DOWN

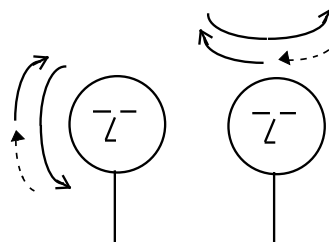


Figure 2. YES NO

		Mode 1			
		NR	YES	UP	DOWN
Mode 2	NR	NR	YES	UP	DOWN
	YES	YES	YES	NR	NR
	UP	UP	NR	UP	NR
	DOWN	DOWN	NR	NR	DOWN

NR = not recognised

Table 1. The resultant gestures when combining two inputs based on a ‘majority rule’ heuristic.

		Mode 1			
		NR	YES	UP	DOWN
Mode 2	NR	NR	YES	UP	DOWN
	YES	YES	YES	YES	YES
	UP	UP	YES	UP	NR
	DOWN	DOWN	YES	NR	DOWN

NR = not recognised

Table 2. More complex heuristic for combining the multimodal inputs.

User	Condition
PJ1	Athetoid Cerebral Palsy
PJ3	Tetraplegia (from head injury)
PJ4	Muscular Dystrophy
PJ5	Spastic Quadriplegia Cerebral Palsy
PJ6	Athetoid Cerebral Palsy
PJ7	Friedrich's Ataxia
PJ8	Athetoid Cerebral Palsy

Table 3. Summary of users and their conditions.

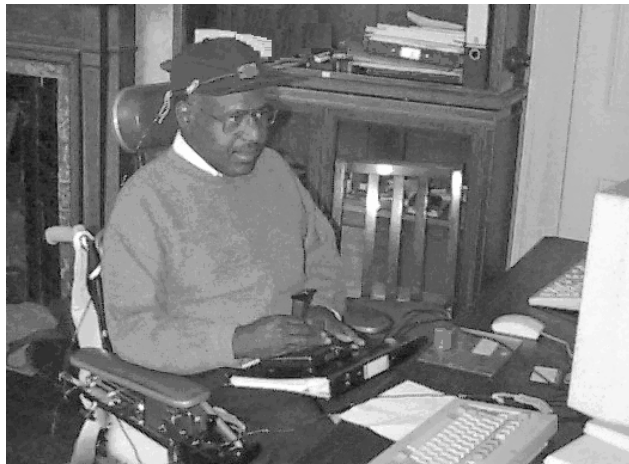


Figure 3. The user trials - this photograph shows both modes of input being used - the Polhemus (on the baseball cap) and the joystick.

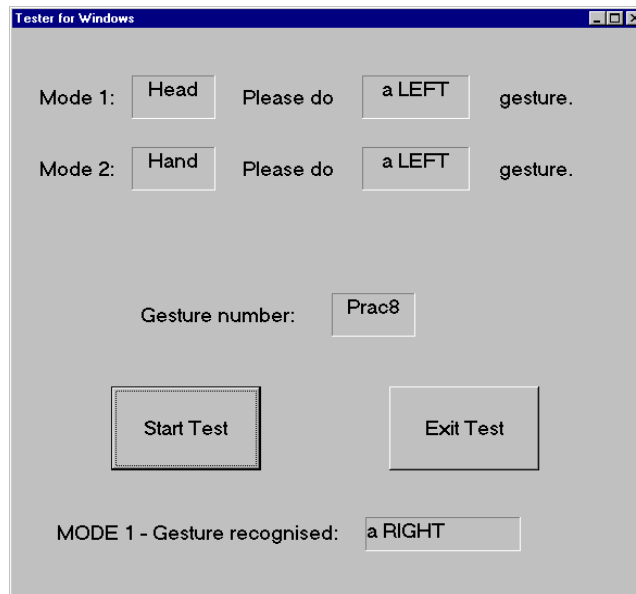


Figure 4. Screen-shot from the Gesture Input Test

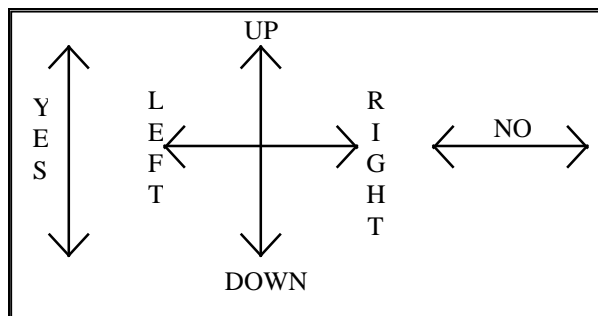


Figure 5. The visual paper cue, placed at the base of the monitor.

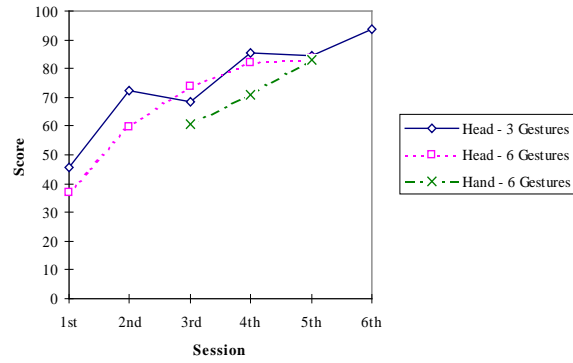


Figure 6. Single mode input and the effects of vocabulary size. This graph shows the scores averaged across the users for each session.

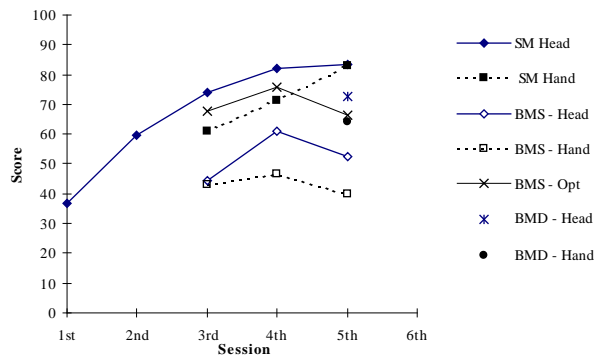


Figure 7. The scores generated by the different input modes during the Gesture Input Test.

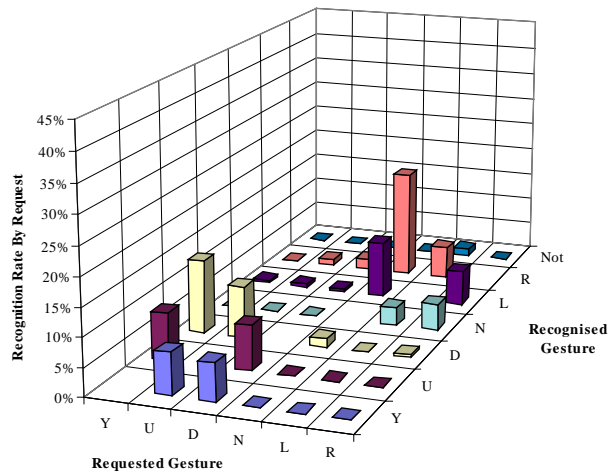


Figure 8(a). The inter-gesture error distribution for head input under Single Mode entry.

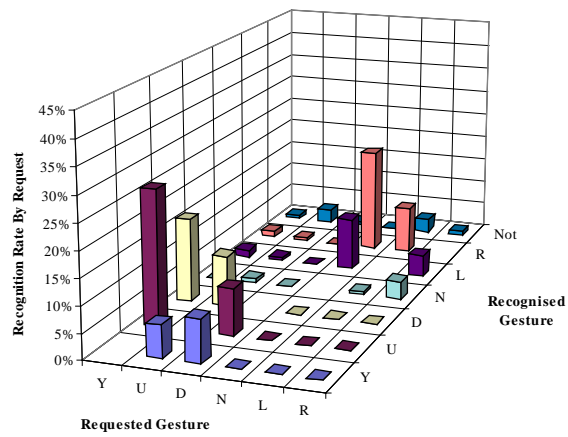


Figure 8(b). The inter-gesture error distribution for head input under Both Modes Same entry.

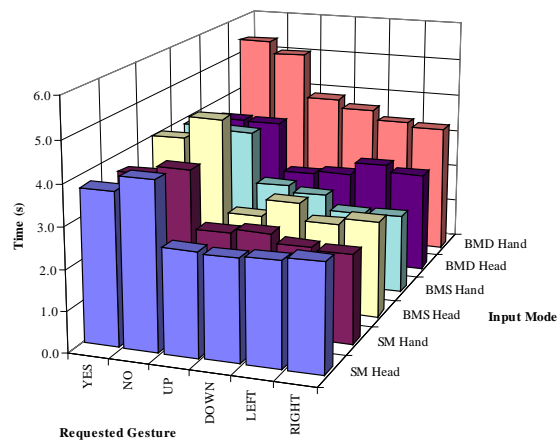


Figure 9. The total gesture times recorded - all users averaged across all sessions.