Int. J. Man-Machine Studies (1986) **25**, 249-269

Touch-sensitive screens: the technologies and their application

J. A. PICKERING

Man-Machine Systems Research Group, Department of Electronics and Computer Science, Southampton University, Southampton, SO9 5NH, U.K.

(*Received 22 October 1985*)

Touch sensing on the surface of a computer display screen is a means of capturing man's natural pointing instincts and using it as mode of human-computer communication. The "touch-sensitive screen" is gradually becoming more popular and it is vital that the individual characteristics of the technologies are understood in order to produce a successful interface. This review covers the methods used for touch sensing on displays, their modes of use and how well they might meet application designers' and users' expectations. None of the technologies available at present are "ideal", particularly for the casual computer user, though they still can provide very successful interfaces if their particular properties are fully taken into account in the target application.

1. Introduction

Pointing is such a natural means of human communication that its potential as a mode of human-computer interaction should be exploited to the full. A human points instinctively by orienting a finger towards the target and touching it, if possible, in order to minimize ambiguity. Many existing computer applications, particularly those involving graphics, require similar pointing action. Such interaction is often mediated through the use of special devices, such as light-pens or "mice", which less directly capture the users' natural pointing instincts. Users will quite unconsciously demonstrate their preference for pointing with a finger when conferring with colleagues about information on a computer display. Displays which are touch-sensitive enable the computer to sense and use this form of communication directly.

Techniques for touch sensing on screens have been available for nearly 20 years yet have not achieved very wide use, despite their seemingly obvious appeal. "Touch-sensitive" is a broad term that is applied to a wide spectrum of devices employing widely differing sensing techniques. A lack of appreciation of the peculiarities of these techniques can lead to disappointing human-computer interfaces. This review will discuss in detail the individual characteristics of the various current techniques and their use and seeks to answer some of the issues raised in the prior review by Pfauth & Priest (1981). The structure of the review centres around a model of the modes of inputs required by computer applications as expressed by so-called "abstract" devices which feature in several graphics software standards. The real devices are then discussed and then their usefulness is then compared and contrasted by means of examining how well each can meet the needs of the abstract devices.

0020-7373/86/090249+21$03.00/0 © Academic Press Inc. (London) Limited

AMP et al.
EUNSUNG-1014

2. The input interface requirements

The computer applications designer is faced with a plethora of interaction devices each of which may be used in a variety of ways to achieve the desired goal. For example, if a mouse is the chosen device, then it may be used:

(a) to pick items from a menu;
(b) position a graphical element or
(c) slide a point up and down a bar to input a numerical value.

These methods of use exploit the ability of the device, together with supporting software, to simulate higher level, "abstract" devices. These are identified as "logical input devices" in a number of graphics interface standards, notably ISO (1983). Logical or abstract devices are introduced in these standards as a mechanism of insulating applications from site dependencies arising from the use of particular physical devices. The graphics environment needs a richer and more demanding user interface and has been the focus of standards embodying abstract devices. Insomuch as office and data-processing applications have increasingly similar requirements, these standards provide a quite general basis for categorizing input tasks. The standards themselves are not as general as they could be, so here the types of abstract inputs identified by Foley, Wallace & Chan (1984) are enumerated:

Select. The user is able to select a chosen item from a set of alternative items. One of the most common types of interaction which maps directly onto this category of input is the menu, where the user selects one from the items displayed as a menu list.

Text. The user is able to specify a symbol sequence. This type of interaction is similar to the select interaction above except that the sequence is composed of individual selections from a given symbol set. Familiar examples are composing a filename or the contents of a file.

Position. The user is able to specify a point in an applications defined space. An example of this type of interaction would be specifying the end-point of a line in a drafting application.

Path. The user is able to express a curve in an applications defined space. This type of interaction is closely related to the position category since, in practice, a path is expressed as a sequence of points implicitly joined by chords. Drawing the curve on a graph in a drafting application would be an example of such an interaction.

Quantify. The user is able to specify a quantity. A simple example of this type of interaction would be controlling the zoom factor whilst inspecting a CAD drawing. The control would enable the drawing magnification to be continuously varied between some given limits.

Orient. The user is able to specify a polar quantity in an applications defined space. This type of interaction is closely related to the quantify category except that quantity being controlled is intrinsically rotational. Many controls such as those in an aircraft simulator require this type of interaction.

The suitability of a physical device as a means of fulfilling the requirements can be judged primarily by the pleasure derived in its use. If a user enjoys interaction with a computer then the relationship will continue and develop. The factors that contribute

to pleasurable interfaces are complex and extend from the physical aspects of the device through to the behaviour of the supporting software. Unsatisfactory interfaces can arise from such compromises as minimizing the number of physical devices available, either for practical or commercial reasons. The objective must be to maximize the overall effectiveness of the complete interface, taking into account the mix of input requirements and the suitability of the available physical devices.

When implementing an abstract device a number of transforming "layers" between the application requirements and the physical hardware have to be constructed. The physical device which the human manipulates comprises the first layer. To minimize the number of layers, enhancing the "immediacy" of the device and its appropriateness, the physical device must match the required properties of the abstract device. Selection and text devices are characterized by the form of the input being discrete. The others are continuous with the orient device also embodying the concept of rotation. There is also a difference between the dimensionality of the continuous devices with both the position and pathing devices implying a sense of space. Thus intrinsically discrete devices, such as keyboards, are not suited to continuous or spacial forms of input but a continuous device can be readily "degraded" to fulfil the requirements of discrete input. Similarly, a potentiometer would not make a suitable position device though a digitizing tablet would suffice for a quantifier.

The inadequacies of the hardware can always be made up for by "thicker" layers with consequent higher cognitive load and poorer performance. Extreme illustrations of multi-layered interfaces are found in computer-based aids for physically less able persons where, for example, a scanning matrix is used to simulate a keyboard. The extra layers are introduced for purely practical reasons so that the user's physical device, such as a sip-puff transducer, can produce the same form of input that an able bodied person would with a conventional keyboard. Some equally extreme but less justifiable examples of "thick" interfaces are often found amongst modern domestic products!

An immediate and natural interface is one which employs the minimum of hardware and software layers. Touch-sensing systems provide a means of implementing the first layer of an abstract device. In order to explore how well each system is matched to the types of input tasks required we need to appreciate the physical properties of the devices themselves.

3. Touch screen device technologies

3.1. GENERAL OBJECTIVES

All the touch-sensitive sensing methods are concerned with the detection of an object on a surface and determining its position. The sensing surface is aligned with another display surface onto which images can be projected. The primary specification for a touch-sensitive system can be summarized:

(a) An uninstrumented finger should be sufficient to use the system.
(b) The sensing mechanism should not obscure or detract from the quality of the displayed images.
(c) The selection of an object and the means of its manipulation should be obvious and "natural".

The most natural means of using a touch-sensitive screen is similar to a keyboard. Two motor actions are involved in making a selection once the target is known and visually captured:

(a) "selection", performed by homing the finger onto the target and then;
(b) the "confirmation" of the selection.

Confirmation is naturally performed by applying an "actuation" force with the finger, the system responding with some combination of visual, auditory and tactile feedback. There are other modes of use of a touch screen which will be covered later.

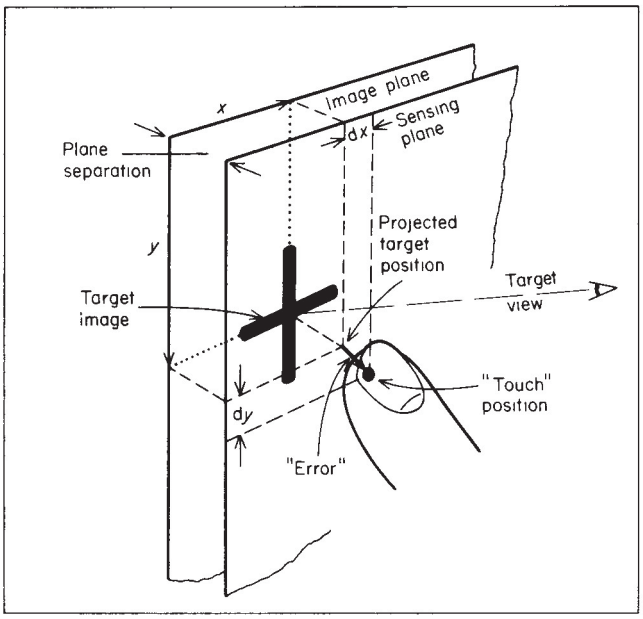


FIG. 1. Touch-screen model.

Figure 1 shows a general model of a touch-sensitive screen with planar image and sensing surfaces. Either the sensing plane or the image plane will be represented by a physical surface from which the user will get a reaction when there is contact with a finger or some other held object. This is the only natural feedback that the device offers and it is at this instant of contact that the user would expect the system to register a "touch". Perturbations between the finger position and the image position will always arise. These contribute to an error vector which may be defined as the difference between the displayed object position and its position as determined by a "touch". Some of the components of the vector are due to the imperfections of the sensing system, namely:

(a) the parallax caused by the plane separation;
(b) the finite resolution of the transducer;
(c) misregistration of the planes and;
(d) transducer drift or failure.

The magnitudes of the error components in each axis, dX and dY, enable the absolute accuracy of the system to be expressed as a ratio e.g. $dX : X_{max}$, where X_{max} is the size of the X coordinate range. When relative finger motion is of importance ("delta" mode) then only the resolution of the detection system is relevant, being expressed either as a ratio or as the minimum distance which is detectable.

The majority of CRT displays present an image surface which is not flat which exacerbates the problem of parallax. The various types of flat displays still do not compete with the colour CRT display in terms of cost and appeal and thus the following discussion will concentrate on the use of CRTs. Some systems necessitate the use of a physically present sensing plane which can attenuate and scatter the light from the image plane, as well as contribute to parallax. The issue of display obscuration and clarity was addressed by Beaton, Schulze & Snyder (1983) whose results are the source of most of the light transmission figures quoted below.

3.2. EARLY SYSTEMS

Johnson (1967) described one of the earliest practical touch-sensitive displays, designed for use within U.K. air-traffic control. The system used "touch wires" attached to the lower portion of the monitor and used the unbalancing of a bridge circuit by body capacitance as the touch-detection method. The touch zones comprised the areas above each touch-wire segment and thus appeared in use much like modern "soft keys". Other systems based on wires have been used, though their obvious disadvantage is display obscuration which increases with the number of touch zones.

Hlady (1969) developed acoustic surface wave technology, which had hitherto been used for digitizing tablets, to produce a touch-sensing overlay. The sensing method has the potential of high resolution though its sensitivity to irregularities on the sensing surface cause its performance to degrade during use. Grease, dirt and scratches cause "clutter" in the reflected signal that render discrimination of a touch more difficult.

A conducting glass touch entry system was developed in the early seventies by Litton Data Systems, described by Marson (1971). This started its life as a light-pen replacement where a metal stylus was used to detect the imposed voltage gradients over the transparent tin oxide coating on a circular, planar glass substrate which was placed over a flat CRT. This proved not to be very durable and was improved by replacing the stylus with a mylar sheet with a conducting undersurface stretched over the glass substrate. The point contact was made when the narrow gap between the two surfaces was closed by finger pressure. This product enhancement created the first membrane touch-sensing system.

3.3. CONTEMPORARY TECHNOLOGIES

3.3.1. Infra-red detection systems

Infra-red technology is one of the most widely used methods for touch sensing. It differs from all the other technologies in that the sensing plane is invisible and thus does not affect the quality of the displayed images. The method first became prominent when it was developed for use in the PLATO Computer Aided Learning project at the University of Illinois described by Smith & Sherwood (1976).

Bezel-mounted infra-red emitter and detector pairs are arranged to create a set of orthogonal beams which may be interrupted by the finger. The beams are usually pulsed in order to increase the beam intensity in comparison with the level of ambient

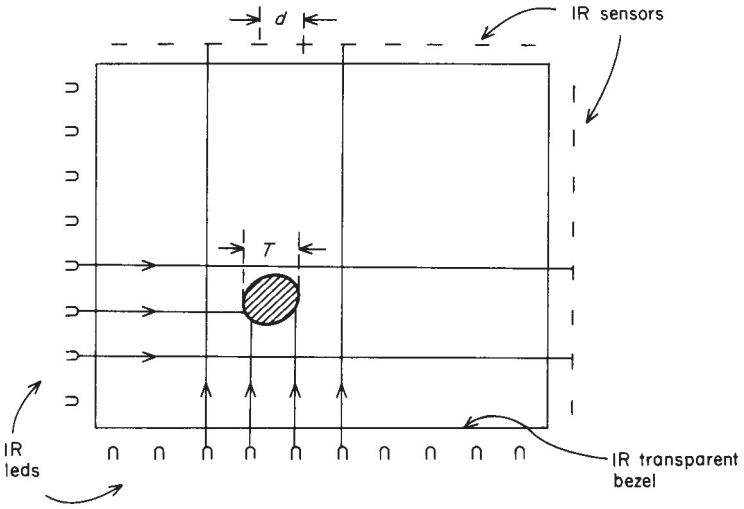


FIG. 2. Infra-red beam-breaking method.

light. Each emitter has only to be visible by its matching sensor on the opposite side of the bezel. This means that if the image surface is cylindrically or spherically curved, as is typically the case with a CRT, the loci of sensors and emitters can also be curved. The beams just graze the surface of the CRT and parallax is minimized at the corners where the problem is worse. The pulsing of the emitters together with the scan of the detectors and the derivation of the touch coordinates is relatively simple and can be done in about 10 ms. The sensor-emitter pitch, d, is chosen such that $T < d < T/2$, where T is a mean finger thickness. Thus a finger will interrupt either one or two beams and with interpolation gives a resolution of $d/2$, independent of the screen size. To achieve higher resolutions in relation to screen coordinates then larger screen sizes can be used with more sensor-emitter pairs in the bezel.

Failure of one of the sensors or emitters does compromise the usability of infra-red beam systems. Problems can also be encountered by stray infra-red from sources such as the sun and artificial lighting. These effects can be lessened by the design of the bezel and the materials used, suitable detector circuitry and the choice of pulse frequency.

Some of the earlier studies of touch-sensitive systems utilized the Marconi "Digilux", or similar, which comprised a large, round, semi-flat air-traffic control style display with 64 sensor-emitter pairs used for each axis, with $d = 6$ mm (Bird, 1977). Systems for retro-fitting to a number of popular visual display units (VDUs) have been marketed for several years by Carroll Touch Technology, Illinois. Hewlett-Packard popularized the method when it became a standard feature of the first version of their HP150 Personal Computer (Straton, McClelland & Kilbourn, 1984).

Higher resolution can be achieved using infra-red sensing if the shadow cast by the finger from each emitter is analysed. The shadow interpolation method, jointly developed by AMTE, Teddington (U.K.) and Glyben Automation (1984) used hardware very similar to that of the beam-breaking system described above. The difference is that, for each axis, the signal from every sensor is examined when each emitter is energized. The sensing surface must be planar because, for each axis, every emitter

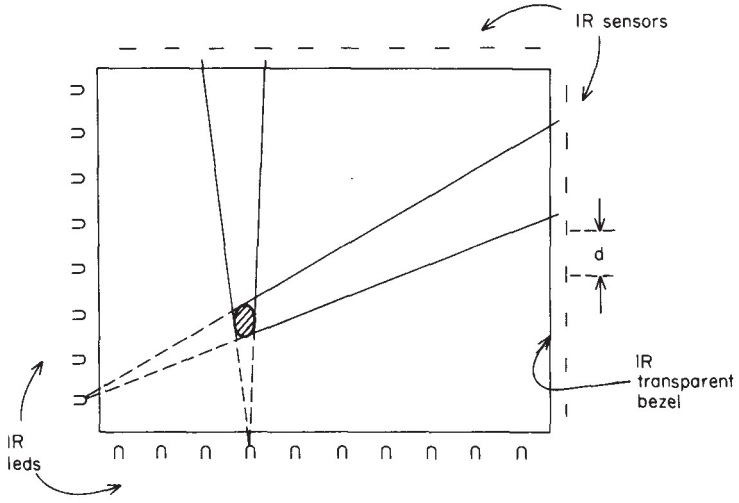


FIG. 3. Glyben infra-red (IR) shadow interpolation.

must be visible by every sensor. This causes its separation from the CRT fascia to rise to 10–20 mm for curved CRTs of only 12–14 in diagonal. Processing is arithmetically intensive and current implementations take about 50 ms to produce a sample, However, the advantage of the method lies in the inherent redundancy of the sensing method. The algorithm averages the effects of all the shadow information gathered and thus is relatively insensitive to individual sensor or emitter failures. Further, because of the large amount of information gathered, the resolution can be better than $d/10$ this is

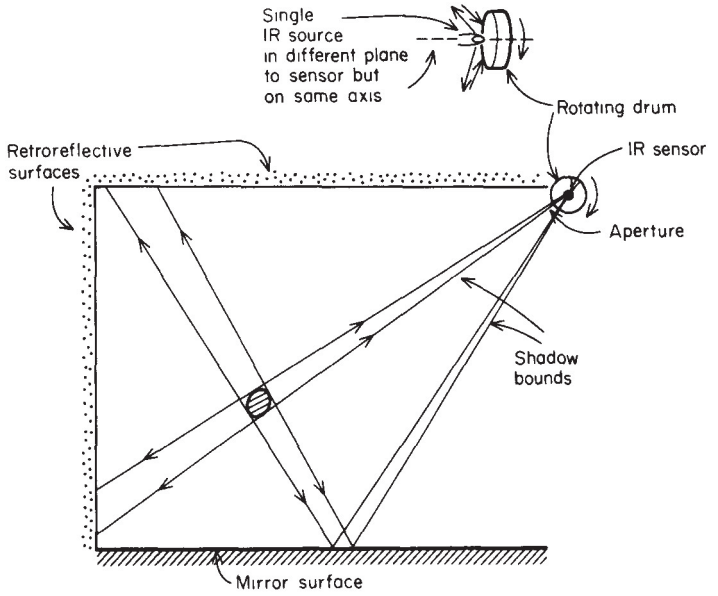


FIG. 4. "Cyclops" infra-red (IR) system.

only slightly affected in the central region for each failure. This with the worst resolution occurring adjacent to the faulty device(s). When failures do occur they can be self-diagnosed whilst the system continues to give service.

Another method that uses shadow information is the "Cyclops" system developed by Illinois Tool Works, Illinois, described by Costlow (1983). The name is apt since sensing is achieved using one light source and one sensor placed in the top corner of the surrounding bezel, collocated in one axis. The source fills the sensing plane with rays both directly and via reflection in the mirror set in the bottom edge of the bezel. These are retro-reflected by material in the top and side bezels back to the sensor adjacent to the source along their originating paths. The angle of receipt of the returning rays is measured by enclosing the sensor in a spinning, slotted drum. When a finger penetrates the sensing plane, the angles of the shadows cast are derived from the sensor and are transformed into the finger position coordinates. In comparison with other infra-red systems it demonstrates a trade-off between circuit and mechanical complexity. However, though the resolution is potentially better than the beam breaking method, it does not compete with the resilience and resolution of the shadow interpolation method. Its sample time is currently slow (100 ms) and it suffers from the disadvantages of a planar sensing surface.

3.3.2. Contact systems

The remaining technologies contrast with the infra-red systems because the measurement methods rely on sensing surfaces that are physically present which can affect the clarity and brightness of the display. The overlay must fit intimately with the display so that the image and sensing surfaces can be as close to each other as possible. Most add typically 5–10 mm to the thickness of the display facia and thus do make parallax effects worse.

The simplest schemes are effectively "translucent keyboards", either using capacitive sensing or cross-point matrices. The zones can be created by fusing transparent conductive pads and tracks into the display surface or by deposition on the surfaces of an overlay. In capacitive systems the pattern complexity restricts the number of touch zones over the surface of a screen to about 20–30. They also suffer from the disadvantage that they require close proximity or physical contact with the body via the finger or some conducting object. The touch cannot be sensed if the user has gloved hands or tries to use an insulating stylus.

The cross-point matrix method utilizes a sandwich in which the inside faces of the outer members have parallel conducting strips set at right angles to one another. The strips are formed by deposition of a thin, transparent conducting film of gold or indium–tin oxide on a mylar substrate. The inside of the sandwich comprises a grid of separators that prevent contact between the outer sheets until finger pressure is applied. The separators are typically small plastic "dots" placed between the strips so that they do not interfere with the contacting required in each zone. The result is much like a membrane keypad except for the use of near transparent materials. Fluke Corporation have used the technique in their 1720A touch display designed for factory automation and testing applications. The small, rectangular monochrome display has a flat surface and the overlay has 60 zones, arranged in 6 rows of 10, designed to coincide with character cell groups. Loss of display clarity is caused by heavily used parts becoming scored and eventually the overlay has to be replaced.

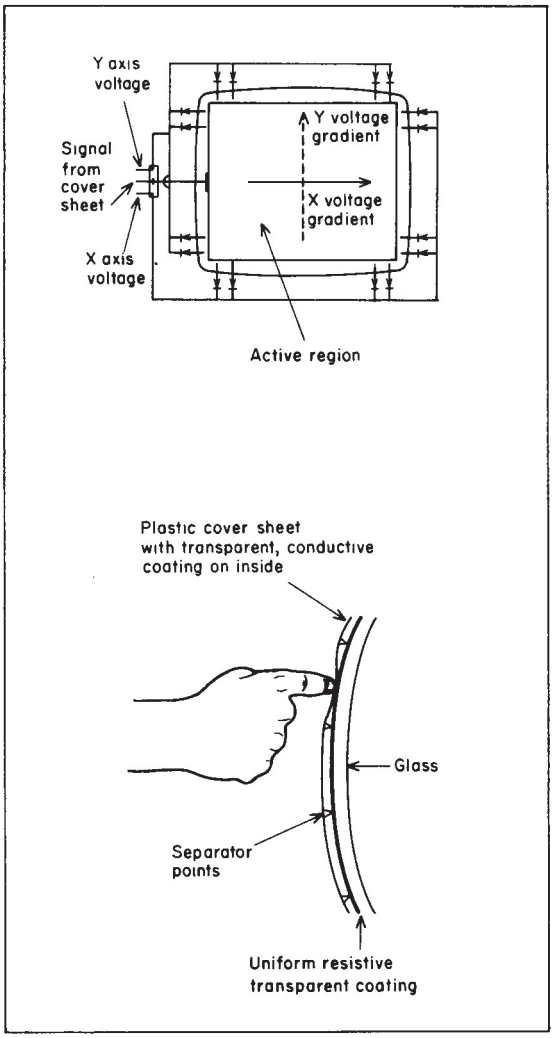


FIG. 5. "Elographics" membrane.

Elographics, Oak Ridge, manufacture a development of Marson's system made up of an overlay sandwich comprising a resistive sheet on a glass substrate and a conducting sheet on a mylar substrate. The sheets are separated by an array of microscopic plastic dots placed at a pitch that enables the sandwich to be made mildly spherical to fit standard CRTs. The light transmission through the assembly is about 50% with the array of separation dots just visible at normal operating distances. The pressure needed for the sheets to come into contact depends on the surface area of the object used to effect a touch. For a finger the pressure is lighter than the actuation force of most keyboards and is proportionally less for sharper objects. The absolute resolution is determined by the quality of the materials and construction and is about 1 part in 100. Relative motion can be detected with resolution determined by the performance of the data converters, which are typically 10-bit devices. The sampling time is determined

by the axis switching time and data conversion rate and is about one sampler per 10 ms. Like the cross-point membrane, the mylar front surface of the overlay is susceptible to wear and damage. Unfortunately, as it is an integral part of a costly assembly it cannot be so readily replaced.

A single resistive sheet is used in a system manufactured by Interaction Systems which utilizes locally applied body capacitance. The overlay comprises a curved glass substrate with a conductive layer fused into the surface. The treatment results in a durable surface but whose light transmission figure is about 60%. The touch coordinates can be resolved to about 1 part in 100 in each axis. As it relies on body contact, it suffers from the shortcomings of capacitive sensing mentioned above.

Developments of the original acoustic ranging overlays are still available (TSD Products), though with moderate resolutions similar to the infra-red beam-breaking technique. The glass is merely used as a surface wave propagation medium and thus has a light-transmission figure of better than 90% and can be curved to fit standard CRTs to minimize parallax. A touch is detected when the finger contact is sufficient to produce a discriminatable echo.

A potentially cheap touch detection method, based on a technique developed by Thompson-CSF (1973), has been developed by Into-Touch, London (U.K.). It uses an untreated, curved glass faceplate mounted onto a CRT using small piezo-electric separators at each corner. The system is called "velocity touch sensing" (VTS) as it is the force impulse caused by the deceleration of the finger as it hits the overlay which enables measurement of its position. The pulses generated by the piezo-electric transducers are measured and transformed into touch coordinates with a resolution of 1 part in 100. Present implementations are somewhat sensitive to the changing static load on the transducers and do not have a wide operating force tolerance. Because of the manner in which the transducers are used the method cannot detect the continuing presence of a finger on the sensing surface.

A feature of all the overlays that use continuous sensing techniques is their inability to distinguish between a single touch and simultaneous multiple touches or the presence of a fist or palm rather than a finger. The systems reply with a centroid coordinate which is indistinguishable to the receiving system from that produced by a single finger. The discrete, transparent keyboard methods can detect multiple touches in the same manner as traditional keyboards because they use similar logic. However, the imaging systems have the potential to identify these types of mistouches as well. Such status information enables the application to ignore the coordinate and/or inform the user how to use the system correctly.

3.3.3. Speculative techniques

The Knowlton (1975) keyboard can be classed as a touch sensing system, though somewhat unconventional. The image and sensing planes are optically combined, using reflection for the image and transmission for the sensing surface. An advantage of this scheme is that the sensing plane does not have to be transparent which enables widely available and cheap sensing technologies to be used such as ordinary keyboard buttons. Its disadvantages are mainly its bulk, restricted field of view and non-obvious method of use.

Work done on touch sensing systems at MIT by Negroponte, Herot & Weinzapfel (1978) investigated the possibilities of a strain gauge-mounted overlay. This method

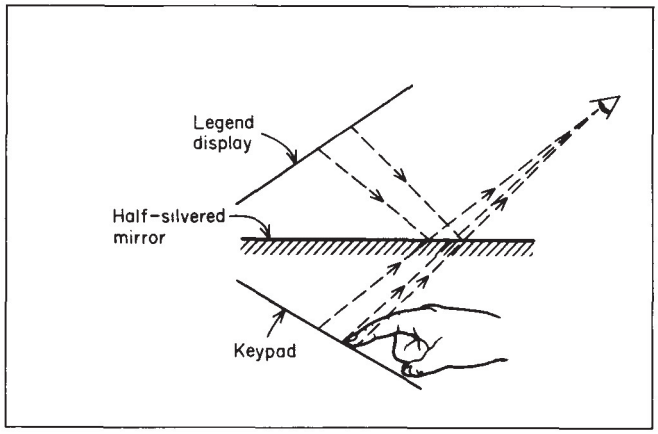


FIG. 6. Knowlton keyboard.

differs from the VTS technique described above because static forces can be measured. Some interesting methods of using a touch system became possible as the magnitude and direction of force exerted by a finger could be measured. In particular, torque forces created by attempting to "twist" the overlay with several fingers can be measured. A number of experiments were conducted to explore the nature and value of these modes of use. These ranged from cursor control using the magnitude and direction of finger force, a duck-shoot where the "bullets" emanated from the user's finger and torque control of a simulated instrument knob. It was hoped that the strain-gauge method alone would enable resolution of finger position but only poor accuracy could be achieved due to sensitivity of the gauges to equipment vibration. Throughout the experiments finger position had to be derived by the commercial acoustic ranging system (Instronics) which constituted the overlay.

An imaging technique utilizing linear charge-couple device (CCD) arrays is described by Funk (1980). Two linear arrays are used at the bottom corners of the screen bezel, each with a slit used as the image forming element, like a one-dimensional

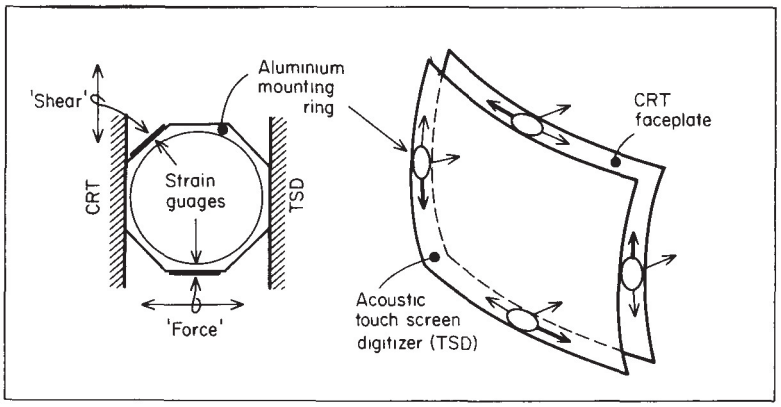


FIG. 7. Negroponte's hybrid employing static force sensing. Example torque sensing vectors in bold.

"pin hole" camera. The sides of the bezel are illuminated so that the CCDs can pick up the reflected light from any object placed in the sensing plane. The coordinates of the object can be found by triangulation from the angle of the object as seen from the two view points.

An optical overlay using the principles of total internal reflection has been described in a patent by Western Electric (1983) and by Schwartz (1986). The body of the overlay consists of a soft, transparent silicone rubber sheet laid over the display. The pressure of a finger locally deforms the sheet and a proportion of the light from the image beneath it is internally reflected. The pulses of light created by the raster scan of the image is detected by sensors coupled to the sheet edge. The finger position is then derived in a similar manner to that used by light-pens.

4. Human factors and performance

4.1. ERGONOMICS

Both the keyboard and the touch-sensitive screen are designed to be operated by fingers. However, intimately associated with the touch device is a display. Since there are differing recommendations for positioning keyboards and screens (e.g. Cakir, Hart & Stewart, 1980) a conflict arises as to where to place a touch-sensitive screen. The conflict was well illustrated by the first release of the Hewlett-Packard HP150 "touch screen" Personal Computer.

The earliest "touch wire" systems took the keyboard approach and inclined the touch display at 30° from the horizontal since at the time that was the recommended angle for keyboards (Hopkin, 1971). Most other systems since have taken the opposite approach and position the display at normal screen height, inclination and distance from the user. An investigation by Beringer & Maxwell (1982) into the effects of touch screen inclination showed no significant performance differences between the conditions tested, 90-, 67-, 45- and 35° from line of sight. However, the subjects did report that use was more fatiguing on the arm and wrist for the extremes of declination, 90- and 35° respectively. This is vindicated by some similar pointing experiments performed by Long, Whitefield & Dennett (1984) where subjects complained of arm fatigue with a 90° display and expressed a wish for an inclined angle. This indicates that a heavily used touch-sensitive screen should be oriented somewhere between the extremes. The shallower inclinations will enhance parallax effects and users may experience difficulties due to their fingers and hand obscuring the target whilst homing. Also coping with reflections and glare from room lighting may become more difficult.

Usher (1983) used a position, compatible with both display and touch operation, which relies on the user being able to rest the elbow on the surface below the display and use it to pivot arm movements. This is satisfactory provided the elbow does not interfere with a keyboard or documents lying on the work surface. Mobility over the screen surface is then more dependent on wrist and finger motions and smaller screens are better matched to this type of use.

4.2. CHOICE OF SENSING METHOD AND ITS USE

The choice of sensing technology has considerable implications for the manner in which the touch system is used and the resulting performance. All the commercially

available systems are supposed to be easy to use because they operate using direct selection "just like a keyboard". However, they all fall short of these user expectations because force-actuated keyboards have a characteristic type of tactile feedback which none of the touch displays can truly simulate. When operating an ordinary keyboard, light physical contact may be used during the selection phase and then pressure is required in the confirmation phase with the user receiving complementary physical feedback. The infra-red and imaging systems are particularly deficient in providing the expected type of feedback since they are not touch-sensitive but "proximity" sensitive. If such techniques are to be used in a convincing keyboard-like way then every effort must be made to minimize the distance between the tangible image and invisible sensing surface. The spherical surfaces of most CRTs make this requirement difficult to meet and contributes to significant additional sources of error that tend to frustrate any attempt at achieving high absolute accuracy.

Many of the other sensing techniques are quite literally "touch" sensitive in that slight contact is required for a touch to be registered. These "light touch" technologies (including infra-red) pose a problem because touches can be spuriously generated during the initial selection phase. The cross-point or resistive membrane technologies are perhaps the most successful at meeting the expectations of the user since a touch is only registered in the confirmation phase when conscious finger pressure is applied.

Indirect selection can be used to overcome the absolute accuracy problems arising from either the user or the sensing system. When a finger activates the sensing surface the application provides feedback by marking the current position with a cursor or highlighting the nearest quantifying object. The user may continue the selection phase by moving the finger until the desired object is marked. Confirmation is then performed by removing the finger from the sensing surface. Operation is rather like using a mouse except that the cursor moves over the screen by following the on-screen finger movement but to effect selection the finger is withdrawn from the screen instead of pressing a mouse button.

As with off-screen position devices, the device motion to cursor movement ratio ("control : displacement" or C : D ratio) can be varied to suit the circumstances. With a touch screen selection can be performed in a two stage process in which the approximate position of the target is obtained from the first coordinate using a 1 : 1 ratio which, for example, then changes to 8 : 1 to locate it more exactly. This technique is peculiar to touch-sensitive screens and light-pen systems and exploits their unique capability of direct selection. The technique enables the effective resolution of a touch screen to be enhanced and in a manner that takes account of the limitations of a finger as an accurate pointing tool. When the target objects are small a biassing transformation between the finger coordinate and the marked position should be introduced to prevent the finger from obscuring the image. This mode of use is particularly well suited to infra-red systems because:

(a) their light touch facilitates quick and easy finger movement and;
(b) the distance between the sensing plane and image surface prevents interruptions in the coordinate stream.

Though indirect selection copes admirably with some of the sensing problems, it does detract from the supposed "*raison d'être*" of a touch-sensitive screen.

4.3. USER PERFORMANCE

All applications using touch-sensitive screens have to perform correlation between the touch coordinates and display coordinates in order to determine the selected item. The various sources that contribute to the error vector complicate this correlation procedure. Some of these are due to the imperfections of the sensing method, such as quantization and misregistration, as mentioned above. However, additional factors due to parallax and finger approach contribute further error components. Their size and direction are due to the sensing system and the particular use of the display and the way the user effects a selection. The remaining components are mainly due to the user and arise from:

(a) eye dominance and handedness;
(b) hit accuracy and;
(c) tendency not wish to obscure the target with the finger.

Hit accuracy is dependent on the reach, screen angle and the state of the user, such as fatigue, which can be improved by suitable screen positioning, as previously discussed.

The studies of Beringer & Peterson (1984) have successfully isolated and quantified some of the error components. The sensing system employed in their studies used infra-red beams at a pitch (d) of 6·35 mm over a 12-in display which precluded very accurate examination of the biasses introduced by users. One of the issues they wished to address was whether some of the sources of error could be reduced by adaptation, either automatically by the system or by user training aided by accuracy feedback. Analysis showed that individual user profiles could be developed and consistently applied to reduce the user error components. A general "population" model could not be usefully developed as factors such as handedness contribute almost equal and opposite biasses. Training seemed to be successful but care is needed to ensure that the user is not coerced to behave unnaturally, such as requiring the target to be obscured by the finger.

If direct selection is to be used and no dynamic correction is to be applied then all components of the error vector must be allowed for in the correlation to reduce the risk of a mistouch. This has the effect of enlarging the touch zone which bounds the perceived target, be it a symbol or some legend. The zone may be marked on the display to emphasize that touches in this zone will be correlated with the enclosed target. The zones may be further separated by guard regions within which touches cause no action. The most consistent error component across the population for which correction can be applied is that of touching "low", the amount of correction being dependent on the size of target in comparison with the finger. Also some correction dependent on the target position on the screen may be applied to allow for the effects of parallax. In its crudest form this would mean increasing the target bounds near the corners of the display where parallax is worst. Almost all the other sources are probably best modelled as random and a suitable allowance made for them. An air-traffic control application described by Gaertner & Holzhausen (1980) reported that the number of touch errors with a special continuous IR system stabilized with touch zones greater than 22 mm in diameter.

The mechanism of indirect selection can nullify most of the sources of error which affect direct selection, but additional sources do arise. The main one is the "fallout" error caused by the finger not being withdrawn vertically from the surface, but coming

away obliquely and dragging through other zones (Whitfield, Ball & Bird, 1983). The infra-red sensing method is particularly prone to this type of error due to the distance between the sensing and image surfaces. This can be accommodated in the usual way by enlarging the touch zones and/or guard zones. Another method is to determine the dwell position of the finger prior to its withdrawal from the sensing plane. This is typically implemented by storing the coordinate stream in a queue and retrieving the coordinate sampled a fixed time interval back from the last detected touch.

4.4. DISPLAY TECHNIQUES

Another factor which influences the effectiveness of an application using touch-sensitive screens is the way in which the display itself is used. Issues such as the size and number of targets on the display will be determined by the performance criteria already discussed. Many applications will use menu-based dialogues and the wider issues of presentation and the structure of the hierarchy will need to be addressed. Snowberry, Parkinson & Sisson (1983) provide some useful information on use of menus and Carter (1982) and Murch (1984) discuss factors affecting the use of colour.

One of the supposed appeals of a touch screen is its flexibility in presentation of alternatives and its scope for being able to adjust these dynamically to suit the context and the user. Gestures over a touch screen become learned in a similar manner to often used patterns on an ordinary keyboard. So, as far as is possible, spatial consistency should be preserved to avoid conflict between the users' spatial memory and the application.

4.5. FEEDBACK AND RESPONSE TIME

The response time of the system is vital to the success of the application. This falls into two categories:

(a) the time to produce the feedback echo, such as moving a cursor, brightening an object or producing a "beep" and;
(b) the time to respond to a command and/or commence a new display.

The former must be created instantly as far as the user is concerned and is an aspect of a touch system that considerably affects user attitude and performance (Hopkin, 1977). Boyle & Lanzetta (1984) provide some figures for the mean time threshold of detection between a user initiated event and the consequent machine feedback, being about 150 ms with about 50 ms being the lowest threshold. Thus the sensing technology must be chosen to provide samples at least as regularly as this with an allowance for generating the echo. Local intelligence near the point of interaction makes this readily achievable. The issues of gross response time are beyond the scope of this paper and the reader is referred to the literature (e.g. Schneiderman, 1984).

5. Further experiences

Whitfield, Ball & Bird (1983) compared the abilities of the "Digilux" infra-red system, a miniature console-mounted digitized pad (the RSRE Touchpad) and a tracker-ball for selection tasks of varying precision in air-traffic control type tasks. All the techniques involve indirect selection which, though not a "natural" method, was appropriate to the skilled operators for whom the devices were designed. The touch-sensitive screen was most liked for all tasks except for those requiring high resolution, where the

tracker-ball was preferred and caused fewer errors. However, selection did take longer with the tracker-ball.

Usher (1983) has conducted studies into the use of infra-red touch systems (Carroll) for plant control in an electricity-generating station control room environment. The device was used to simulate a selection device, using direct selection, and a quantifier for control adjustment. An alternative system used for comparison had a numeric keypad for both selection and quantification. Performance was not appreciably different between the two systems, though the users expressed a significant preference for the touch screen version. Mistouches in light touch systems is a problem noted in the earliest systems and is worse with proximity sensing technologies. Thus in this application particular attention had been paid to making the dialogue tolerant of mistouches. Any action that had irreversible consequences was always checked by an explicit confirm or abort phase.

For a number of years EASAMS Ltd (U.K.) (1983) have employed infra-red beam sensing systems (Carroll) as input devices in their general-purpose EAMACS interactive interface. The system has been successfully used in a number of projects, both civil and military, concerned with graphical information retrieval, manipulation and control. The interface features simulations of all abstract input devices, though paths are specified as position sequences and orientation is imperfectly simulated. The 20-in displays used enable the best resolution to be achieved with the sensing technique, though at the cost of larger corner parallax. This is coped with by using indirect selection throughout and quite large "button" touch zones, with generous guard zones, arranged in a fixed 8×8 matrix. The display size and zone size facilitate use of a touch screen by other persons sitting or standing around the operator. For selection of small objects and positioning the technique of changing the C:D ratio is used and the effective resolution can be still further enhanced, if necessary, by zooming the display.

Schulze, Beaton & Snyder (1983) have studied the differences between various touch-screen technologies in three model applications that utilized direct selection. Marked error performance differences between the methods which, though not very consistent between the applications, give an indication that light-touch technologies give rise to higher error rates than force-actuated systems.

Karat, McDonald & Anderson (1984) have conducted trials to compare both performance and user preferences between a touch screen (Elographics), keyboard and mouse. The experimental applications were a telephone directory aid and appointments diary and involved selection and, in some of the tasks, some text input using a standard computer keyboard. There was a strong user preference for the touch-screen version of the models which was only moderated when tasks also involved text input. In all cases the mouse was least liked and, despite extended practice sessions, was the slowest method. These comprised simple "target practice" in which the participants were asked to make selections as fast as possible and were not penalized for errors. Analysis of their accuracy showed that the touch screen and mouse were significantly more error prone than the keyboard. The targets were boxes measuring only 14×13 mm which may have contributed to the touch-screen error rate.

It may be fairly concluded that touch screens are generally liked by users, provided they are given the time and opportunity to get familiar with them. Even so, they seem to have a higher associated error rate than the more robust and traditional keyboard

interfaces. Thus the dialogue must be particularly "gentle" with users in order not to put them off during the initial familiarization period.

6. Abstract device mapping

The survey of the available device technologies and experiences may now be summarized by assessing how the capabilities of the real devices map onto the requirements of abstract input devices.

Selection. The implicit motive behind the development of touch-sensitive screens was to create a flexible, application independent selection device. From the above discussion it appears that this has been reasonably successful. All the techniques are able to display images whose variety and detail is mainly limited by the sophistication of the display controller and screen.

When employing direct selection the limit to the number of target items is determined by the finger's ability as a pointing instrument in the screen environment. Coarse resolution detection systems, such as the discrete methods, confine the size, number and distribution of targets to that of the detection zones. Techniques featuring higher absolute resolutions give the applications designer greater freedom in the layout and design of the targets. However, allowances must be made for the error components that arise due to the users' behaviour during interaction and these can become dominant with resolutions beyond that of a finger thickness, particularly with infra-red sensing systems. The sampling time of the device must be such as to give the user the impression of instant feedback and allow for key "stabbing" where the dwell time of a finger in a touch zone can be less than 100 ms.

Perhaps the most important factor when using direct selection is the ability of the technology to match the keyboarding metaphor. Currently the membrane technologies are the closest to meeting this requirement. Unfortunately, they are the most susceptible to damage in the public environments where the metaphor is most appropriate.

If the target sizes become too small for acceptable selection directly then indirect selection must be used. The touch technology must be capable of generating coordinate streams from finger movements and at sufficient pace to allow the feedback echo to appear to follow finger movements. Low relative resolution systems can be employed by changing the C:D ratio once coordinate streaming commences, though it must be sufficient to allow for reasonable initial accuracy and minimal movement for final homing. The infra-red technologies are quite well suited to this mode of use as little or no pressure is required to generate coordinates, thus facilitating finger movement. This type of indirect selection works in an almost opposite fashion to keyboarding and thus requires some familiarization and practice.

Text. Text involves sequential selection from an essentially unchanging choice. Changes to the display to reflect contextual restrictions should be minimized to avoid distracting the user and allow anticipation of the next selections. All the comments made regarding the simulation of a selection device apply to a text device but with the following qualifications. Since the display will be more or less

static while the text sequence is built, selection will tend to take place faster with the minimum of finger and arm movement. The direction of finger approach is more variable and this changes the nature of the error components that should be taken into account when designing the size and layout of touch zones and guard regions. There is a greater risk of contact with the physical surface during the selection phase which militates against proximity and light-touch technologies. Even the membrane technologies are prone to mistouches if stroked by a finger nail.

Position. The method of indirect selection above demonstrates the method by which a two-dimensional position of graphics precision may be input. A high relative resolution sensing system is appropriate since it more nearly matches the continuous requirement of this type of device. The coarser resolutions may be used but at the cost of introducing higher C:D ratios and the extra finger movement required to achieve the target position. The final accuracy of the position will be determined by error components such as "fallout" and sensor jitter, though the significance of these can be reduced in proportion to the C:D ratio in force.

It is questionable whether a touch-sensitive device is suited to positioning tasks as the finger is not expected to be a precise pointing tool. However, as demonstrated in two of the reported studies, a touch screen with limited resolution can still be potentially faster than conventional positioning devices because of its direct selection capability. Positioning is more naturally achieved with a "sharper" tool such as a stylus which, given the choice, a user may prefer to use in such circumstances anyway. Both these aspects are met in light-pen systems which do have the potential for high resolution.

Path. Pathing is a task which is more familiarly performed with the aid of an instrument because the action is so similar to drawing with a pen or pencil. Thus again, it is perhaps not appropriate to use a finger for this type of input. If it is to be used then the sensing technology must be of sufficient precision at a 1:1 C:D ratio since any other alternative, such as identifying each position in the path sequence, is not satisfactory. Also it must be able to generate a coordinate stream at a sufficient rate to respond to a finger stroke. The resistive membrane is the method that most closely meets the requirement of high relative resolution and short sampling time but the necessary finger pressure may make stroking awkward. Use of some other instrument, like a biro, would not be encouraged as damage to the overlay may quickly result. The infra-red shadow interpolation method may be more suitable but does not generate coordinates fast enough and with sufficient precision because of jitter and fallout errors.

Quantify. The quantification task is similar to that of positioning, except that only a single one-dimensional value is required. It is usual to show the full range of the value to be quantified as some form of bounded line or bar accompanied by the current instantaneous value. A finger placed on the bar and moved up and down will cause the value to change. If it is required to be able to move quickly over the full range of the value and have fairly fine resolution while doing so, then a high resolution, light touch-sensing technology is required. If the value input by the user is usually initially about right then coarser resolution, light touch methods can be used, using a change in C:D ratio. All the current technologies are limited to working with a single finger and thus several quantifiers cannot be operated simultaneously like the controls on a sound mixing desk.

Orient. The implicit rotational requirement of this type of device cannot be directly simulated by any currently available touch technologies. Only the strain gauge detection technique can produce a satisfactory simulation. Thus the usual alternative employing a quantifier must suffice.

A summary of how well the device mappings are achieved between the various current technologies is given in Table 1. As can be seen, there is no "ideal" touch sensing technology that covers all requirements.

TABLE 1
Device mapping summary

Abstract device \ Technology	Discrete capacitive	Infra-red beam breaking	Continuous capacitive	Infra-red shadow interpolation	Discrete membrane	Continuous membrane	Acoustic	Piezo VTS
Selection	○	○	○		●	●	●	●
Text	○		○	●	●	●	●	○
Position		○	○	●		○		
Path				●		○		
Quantify				●		○		

Key: ●○ ← BEST

7. Conclusion

The survey of "touch screen" technology and experience seems to show that the devices currently available do not entirely match with the users' expectations. The truly touch-sensitive screen that behaves like a keyboard and is robust as a keyboard is currently unavailable, despite marketing opinion to the contrary. However, the devices can be useful in their own right if used appropriately. The designer must analyse the interaction requirements of the application and judge whether a "touch screen" is really appropriate, taking into consideration the individual performance of the technologies and the mix of interaction requirements. The issue of cost has not been addressed in this review and this must also be considered in many applications. Touch-sensing systems are currently quite expensive in small quantities, typically doubling the price of a standard monitor or terminal.

All the commercially available techniques have concentrated upon one particular sensing method and have optimized it to meet the detection requirements. There are

good commercial reasons for this as the market is strongly oriented towards the cheapest devices that are easily retro-fittable to existing displays. Some of the problems may be ameliorated by incorporating suitable user models in the system software, either in the form of user profiles or by means of dynamic adaptation during use. However, no single technique can meet the users' natural expectations regarding a keyboard analogue and improvements are needed here as well. It might be that hybrid systems, like that of Negroponte *et al.*, may provide more satisfactory solutions.

The author wishes to thank the many people who have shared their experiences with touch systems, either verbally or in their publications. Acknowledgement is given to the U.K. Science and Engineering Research Council for supporting aspects of this work under Grant GR/D/59229.

References

BEATON, R. J., SCHULZE, L. J. H. & SNYDER, H. L. (1983). Touch entry devices and human performance. Part I: Quantifying the display quality. *SID 83 Digest*, Coral Cables, 1983, pp. 162–163.

BERINGER, D. B. & MAXWELL, S. R. (1982). The use of touch-sensitive human-computer interfaces: behavioural design implications. *Proceedings of the Human Factors Society, 26th Annual Meeting*, 1982, p. 435.

BERINGER, D. B. & PETERSON, J. G. (1984). Underlying behavioural parameters of the operation of touch-input devices: biases, models and feedback, *Technical Report HFRL-84-1*, University of Wisconsin-Madison.

BIRD, P. F. (1977). "Digilux" touch sensitive panel. *Proceedings of the IEE Conference on Displays for Man-Machine Systems*, 1977, pp. 28–30.

BOYLE, J. M. & LANZETTA, T. M. (1984). The perception of display delays during single and multiple keystroking. *Proceedings of the Human Factors Society, 28th Annual Meeting*, 1984, pp. 263–266.

CAKIR, A., HART, D. J. & STEWART, T. F. M. (1980). *Visual Display Terminals.* Wiley.

CARTER, R. C. (1982). Visual search with color. *Journal of Experimental Psychology, Human Perception and Performance*, **8**, 127–136.

COSTLOW, T. (1983). Retroreflectors cut component count in CRT touch panels, *Electronic Design (USA)*, 15 September 1983, 42.

EASAMS (1983). *A Description of EAMACS.* EASAMS Ltd, Frimley Road, Frimley, Camberley, U.K.

FOLEY, J. D., WALLACE, V. L. & CHAN, P. (1984). The human factors of computer graphics interaction techniques. *IEE Computer Graphics and Applications* **4**, 13–48.

FUNK, B. K. (1980). CCDs in optical touch panels deliver high resolution. *Electronic Design (USA)*, 27 September 1980, 139–143.

GAERTNER, K. P. & HOLZHAUSEN, K. P. (1980). Controlling air traffic with a touch sensitive screen. *Applied Ergonomics*, **11**, 17–22.

GLYBEN (1984). *Position Detector System.* British Patent No. GB2133537A.

HLADY, A. M. (1969). Touch panel for graphical input: a touch sensitive X-Y position encoder for computer input, *Proceedings of the AFIPS Conference* 1969, pp. 545–551.

HOPKIN, V. D. (1971). The evaluation of touch displays for air traffic control tasks, *Proceedings of the IEE Conference on Displays*, 1971, pp. 83–90.

ISO (1983). *Graphical Kernal System (GKS)—Functional Description.* ISO DIS 7942.

JOHNSON, E. A. (1967). Touch displays, a programmed man-machine interface. *Ergonomics*, **10**, 271–277.

KARAT, J., MCDONALD, J. & ANDERSON, M. (1984). A comparison of selection techniques: Touchpanel, mouse and keyboard. *IBM Technical Report* TR-51.0166, Austin, Texas.

KNOWLTON, K. C. (1975). Virtual pushbuttons as a means of person-computer interaction, *Proceedings of the IEEE Conference on Computer Graphics, Pattern Recognition and Data Structure*, May 1975, pp. 350–351.

LONG, J., WHITEFIELD, A. & DENNETT, J. (1984). The effect of display format on the direct entry of numerical information by pointing. *Human Factors*, **26**, 3–17.

MARSON, R. K. (1971). Conducting glass touch-entry system. *SID 71 Digest*, 88–89.

MURCH, G. M. (1984). Physiological principles for the effective use of color, *IEEE Computer Graphics and Applications*, **4**, 49–54.

NEGROPONTE, N., HEROT, C. & WEINZAPFEL, G. (1978). One-point touch input of vector information for computer displays. *MIT Report NN-100-2*, Massachusetts Institute of Technology.

PFAUTH, M. & PRIEST, J. (1981). Person-computer interfaces using touch screen devices. *Proceedings of the Human Factors Society 25th Annual Meeting*, 1981, pp. 500–504.

SCHULZE, L. J. H., BEATON, R. J. & SNYDER, H. L. (1983). Touch Entry Devices and Human Performance. Part II: Assessing operator performance. *SID 83 Digest*, Coral Cables, 164–165.

SCHNEIDERMAN, B. (1984). Response time and display rate in human performance with computers, *ACM Computing Surveys*, **16**, 265–286.

SCHWARTZ, T. A. (1986). The AT&T soft touch-sensitive screen. *AT&T Technical Journal*, **65**, 62–67.

SMITH, S. G. & SHERWOOD, B. A. (1976). Educational uses of the PLATO computer system, *Science*, **192**, 344–352.

SNOWBERRY, K., PARKINSON, S. R. & SISSON, N. (1983). Computer display menus. *Ergonomics*, **26**, 699–712.

STRATON, P. R., MCCLELLAND, S. R. & KILBOURN, T. E. (1984). The HP150 touchscreen: an interactive user input device for a personal computer. *Hewlett Packard Journal*, August 1984, 11–15.

THOMPSON-CSF (1973). *Position Indicating System.* British Patent No. 1308548.

USHER, D. N. (1983). Comparison of touch-sensitive VDU and computer aided keypad for plant control. *Displays*, **4**, 157–161.

WESTERN ELECTRIC (1983). *Touch Sensitive Device.* British Patent No. GB2074428B.

WHITFIELD, D., BALL, R. G. & BIRD, J. M. (1983). Some comparisons of on-display and off-display touch input devices for interaction with computer generated displays. *Ergonomics*, **26**, 1033–1053.