Hands-Free Input Devices for Wearable Computers

Gloria L. Calhoun and Grant R. McMillan
Human Effectiveness Directorate
Air Force Research Laboratory

Abstract

The advent of wearable computers marks a potential revolution in human-machine interaction and necessitates an expansion of control and display capabilities. Several emerging technologies can provide operators with a variety of new channels for interacting with wearable computers. Enabling technologies use signals from the brain, muscles, voice, lips, head position, eye position, and gestures for the control of computers. These hands-free, head-up controllers may be required to fully exploit the advantages of wearable computers. This paper describes several hands-free controllers that are candidate input devices, either individually, or as part of a multimodal interface. Controller design, task-controller mapping, and other application issues are also presented.

Introduction

Significant advances in wearable computer technology promise to revolutionize our interaction with information systems. Having tetherless mobile computer and telecommunications systems will allow workers to access information, communicate with others, and store information, all while performing their field operations. This capability will be valuable in many work settings, including rural/home health care, maintenance tasks, and on-the-job field training. Additionally, wearable computers can enrich our personal education and entertainment; picture a student retrieving additional information on items of interest viewed in a museum or a hobbyist accessing supplemental data (e.g., airplane model schematics or needle-point directions).

One key to the successful application of wearable computers is enabling seamless interaction with the computer in the context of performing other tasks. Even though one-handed keyboards [10] and special purpose mice, trackballs, dials [1], and finger sensors [6] have been designed for wearable computers, there are instances where it is preferable for an operator's hands to remain engaged in their primary task. Speech input provides one such alternative, but is constrained by noise and may not be applicable in some task environments. Compact nonconventional controllers, that do not require a direct mechanical linkage between the operator and the input device, are desirable to provide hands-free interfaces that can supplement manual control. The Alternative Control Technology program is currently evaluating the use of signals from the brain, muscles, voice, lips, head position, eye position, and gestures for the control of computers and other devices [3,4,11-13]. These emerging hands-free technologies provide operators with a variety of new channels for controlling wearable computers.

Objective

The purpose of this paper is to provide a starting point for interface designers considering alternative controls for wearable computers. Based on our experience with nonconventional controls, candidate hands-free input modalities for wearable computers are identified and briefly described. Additionally, some constraints that must be considered when applying hands-free controllers in wearable environments are outlined. We hope that sharing our experience to date with these novel controllers will inspire their application to this exciting human-machine interface challenge and will encourage further research and development for specific wearable applications.

Candidate hands-free input modalities

A comprehensive description of nonconventional control technology has been documented by Air Force scientists [11]. The following summary discusses the hands-free alternatives most appropriate for wearable computer applications.

Speech-based control

Speech-based control uses pattern recognition methods to map an input speech waveform into corresponding text or a discrete output. Commercially available speech systems are fairly mature making their application both feasible and cost effective. For applications in which speech commands are used to sequence through menu items, isolated or connected word systems will suffice. Changes in the computer display (e.g., icon highlighting) can serve as feedback to the operator that the voice command was recognized. Continuous speech systems that do not require a pause between words are more appropriate for applications in which the operator uses the system to fill out information fields, report forms, etc. [9]. In this application, a transcription of the speech should be available for presentation if the operator wants to monitor the accuracy of the entries.

Despite optimization techniques that can make speech-based control more robust, a key issue is whether speech systems can perform in high and dynamic noise environments [8]. At a minimum, the microphone should be positioned close to the mouth, which complicates operators' hardware. Another challenge is efficient dialogue design such that the vocabulary and syntax are manageable, without imposing a great memory load on the operator. Speech-based control must also not interfere with operator communications and making verbal inputs, that can be heard by others nearby, must be acceptable in the user's environment.

Eye-based control

For applications in which the operator views a display during control operations, harnessing the direction of eye gaze promises to be a very natural and efficient control interface [4]. Careful interface design will be necessary, though, to ensure that the operator's eye movements during task completion are natural and not fatiguing. Thus, a fully "eye-driven" mouse is not envisioned. Rather, natural eye movements should be used to provide a direct pointing capability that can be combined with other hands-free control modalities to command activation. The task design also needs to take into account the accuracy limits of the eye-tracker. For many proceduralized tasks, some intelligence can be built into the system to help interpret the eye inputs, e.g., given previous steps and eye movement patterns, predict the functions that are most likely to be activated next.

With a head-mounted display, head tracking is not required and the calculation of eye gaze is simplified. Methods for tracking the eye can be classified into those that: 1) measure the electric potential of the skin around the eyes, 2) involve image processing of one or more features that can be optically detected on the eye, and 3) employ a special contact lens that facilitates eye position tracking [3]. The latter method is too intrusive for the targeted application.

The first method is probably the least expensive and easiest to implement. Electrooculography (EOG) is based on the existence of an electrostatic field that rotates with the eye. By recording small differences in the skin potential around the eye with electrodes, the position of the eye can be detected. Ideal locations for the electrodes are the upper and lower lids for detecting vertical movements and on the external canthi for horizontal movements. It is unlikely, however, that accurate point-of-gaze tracking is feasible outside the laboratory due to many factors that can result in nonlinear output functions and significant dc drift [3]. For instance, skin resistance varies over time and the corneal-retinal potential itself varies with light adaptation, alertness, and diurnal cycle. Muscle action potentials or external electrical activity can also produce interference. The drift inherent in EOG measurements makes this technology more suitable for measuring eye velocity and acceleration profiles rather than measuring eye point-of-gaze. However, it is possible that EOG tracking would suffice if the control operations only required detecting whether the operator is looking generally left, right, up, or down.

The most practical line-of-sight measurement techniques involve image processing of one or more features that can be optically detected on the eye. Typically, these features are reflections from an infrared source directed at the eye [3]. Methods that involve comparing the reflection relationships between the corneal and pupil or the corneal and 4th Purkinje image (reflection from the rear surface of the eye lens) are the most accurate. However, these methods, besides being more complex and costly to implement, are very sensitive to changes in ambient illumination level and placement of tracking components. Both limbus and pupil tracking methods are alternative solutions for head-mounted displays. In limbus tracking, the boundary between the white sclera and the iris of the eye is detected relative to the head. Since the vertical movements of the limbus are covered by the eye lids, this method is only practical for tracking horizontal eye movements. Pupil tracking is similar to limbus tracking, only the smaller boundary between the pupil and the iris is used. Since the border of the pupil is often sharper than that of the limbus, improvements in resolution can be realized. In addition, the pupil is less covered by the eye lids, enabling vertical tracking.

A light source and photocell, mounted on a "glasses" frame, can also detect whether an eye is open, closed or anywhere in between (an open eye reflects less light than a closed eye) [7]. Although originally designed to monitor the alertness of an operator, this method can also enable a purposeful blink to serve as a control input signal. The usefulness of "blink control" will have to be weighed in comparison to its interruption in viewing the display and its naturalness in use.

Gesture-based control

There are a variety of static and dynamic signs that have been referred to as "gestures," including: "body language," hand/finger forms, grasp of open space, and involuntary motions. Likewise, there are a variety of techniques to read hand and body movements directly [16]. Since the body and hands can be involved in other activities, gesture-based control may best involve detecting defined movements or positions of the operator's face or lips. Optical, magnetic and ultrasonic sensing technologies have been used to monitor an operator's mouth movement. In one implementation, a headset boom located in front of the speaker's lips contains an ultrasonic signal transmitter and receiver. A piezoelectric material and a 40 KHz oscillator are used to create a continuous wave ultrasonic signal [8]. The transmitted signal is reflected off the speaker's mouth, creating a standing wave that changes with movements in the speaker's lips. The magnitude of the received signal is processed to produce a low frequency output signal that can be analyzed to produce lip motion templates.

There are two candidate applications of lip motion measurement. In one, lip movements are processed during speech inputs to provide "lip reading." An experiment using an ultrasonic lip motion detector in a speaker dependent, isolated word recognition task demonstrated that the combination of ultrasonic and acoustic recognizers enhances speech recognition in noisy environments [8]. Alternatively, symbolic lip gestures can be translated into communication tokens that are used as control inputs. Lip gestures do not have high resolution and cannot be used for tasks requiring precise control. They need to be concise and quickly delivered, in order to minimize fatigue and interference with speech communication. Moreover, operators must find the gestures acceptable to employ around others.

There are other facial signals that can be detected, although much of this technology is still immature. An infrared source and associated detector that illuminates the underside of the chin can be used to detect tongue position. By employing special mouth pieces, tongue-operated pointing [15] and tongue-operated keypads have been demonstrated [14]. It is also possible that unique acoustic signatures can be created by different teeth clicks that, in turn, can be harnessed for control.

Electromyographic (EMG)-based control

EMG-based control uses the electrical signals that accompany muscle contractions, rather than the movement produced by these contractions, for control. Electrodes positioned on the surface of skin detect the asynchronous firing of hundreds of groups of muscle fibers. Most commonly, these electrical signals are compared to some threshold value to derive a binary control input – above threshold initiates one control action, below threshold initiates another [13]. Some algorithms employ "time proportional techniques" where the control action continues as long as the operator holds the signal relative to the threshold. True proportional control is difficult, but two to four discrete EMG signal levels can be achieved with training. Continued development is required to optimize the signals employed, assess the stability of the electrode contact over time, and minimize the effect of user movement and external electrical activity on signal recordings.

To implement EMG-based control, it is important to choose a body movement that does not interfere with the operator's normal functions, is not likely to be made during normal task activity, or can be implemented in such a fashion that the system can discriminate a purposeful EMG input from an inadvertent one. For instance, in one concept demonstration developed in our laboratory, operators raise an eye brow or clench the jaw to make control inputs (enter and tab, respectively) for a task presented on a head-mounted display. Semi-dry electrodes integrated into the display assembly detect the changing electrical activity produced by these subtle gestures, and employ these signals to sequence through procedures and graphics.

Electroencephalographic (EEG)-based control

EEG-based control translates the electrical activity of the brain into a control signal for a machine or computer. In one approach, EEG patterns are brought under conscious voluntary control with training and biofeedback [12]. This approach is not appropriate at this time for wearable computers because of the significant training investment. A more applicable approach harnesses naturally occurring brain rhythms, patterns, and responses that correspond to human sensory processing, cognitive activity or motor control. One example is the "P300" brain response that varies with stimulus probability and task relevance. With careful design of the task format and procedures, it is possible to use the natural variance of the P300 for task control [5].

A method we are investigating that involves less impact on task design is based on brain responses to modulated stimuli [12]. These brain responses include components that modulate at the same frequency as the evoking stimuli. Thus, if selectable items of a display are modulated at different frequencies, the operator's choice between selectable items can be identified by detecting which frequency pattern is dominant in the visual evoked brain activity. The operator gazes on the desired selection and the controller registers the corresponding frequency of the displayed item. In preliminary studies, we are using a head band which positions coated plastic electrodes over the occipital cortex. (Aloe vera gel is used to improve contact.) Selection times of 1-2 seconds are achieved by most users. Although detection of these responses is easily accomplished with inexpensive components, optimization of this alternative control requires minimizing the time required for signal processing, developing easily donned electrodes, and minimizing the distraction produced by flashing display items.

Some applications issues

Despite the success demonstrated in laboratory evaluations, these and other hands-free modalities need further evaluation to specifically assess their utility for wearable computer applications. Very few devices are commercially available and, except for speech-based control, none have been configured specifically for wearable applications. However, given the advantages of hands-free control, further research and development are clearly warranted. The following sections discuss some issues that are relevant to the design and implementation of hands-free controllers for wearable computers.

Mobility requirements

Just as wearable computers are designed for mobile operators, input devices must also be mobile. Lightweight, compact, and comfortable components are required and wireless transmitter technology should be employed as much as possible. Ideally the input device should be operable regardless of operator movement or position, and easy to don and doff.

Application environment

Hands-free controllers need to be operable in any environment in which wearable computers are employed. Environmental factors that may impact application include ambient noise, light, temperature, smoke, and industrial contaminants. Some controllers (i.e., speech) may not be appropriate when privacy or covertness is a concern. Likewise, collaborative use of computing and networking technologies can impact input device choice. Since wearable computers are likely to be employed in the presence of others, the input device and operator responses need to be inconspicuous such that there is no interference with ongoing communications and no negative social response.

Target operator population

Implementation of hands-free controllers will be impacted by whether the general population is targeted or whether the controller can be customized for an individual operator. If the former, the design should assume that the operator has little computer sophistication. For the general population, procedures to employ the controller need to be obvious, natural, and require little, if any, training. Likewise, the need for operator calibration and adjustments should be minimized. However, for more specialized military or technical applications it may be acceptable, and even advantageous, to customize the controller and/or utilize a longer training protocol. For example, tailoring signal detection algorithms to an operator's EMG response or training on a speaker dependent speech recognition system may produce significant long-term payoffs.

Task requirements

To date, the control achieved with most hands-free devices can be described as rudimentary. Any application must take into account the limited dimensionality, accuracy, speed, and bandwidth of control afforded by these devices. For some applications and target users (for instance, those with severe physical limitations), speed and accuracy of control may be of less concern, since conventional control options are not possible. Other applications require more rapid and error-free performance. Since hands-free controls do not require associated limb or hand movement, control inputs are typically very rapid, unless lengthy signal processing is required. Rather, it is the precision and accuracy limits of the fundamental human responses and of the controller hardware and software that constrain application. In light of these limitations, efficient procedures are needed for correcting erroneous entries and for safeguarding the system from hazardous control inputs.

The characteristics of the task must also be considered [1]. If the content of the data input is not known in advance, then the input device needs to support arbitrary input. In this case, keyboard surrogates or speech recognizers are likely candidate devices. If the input content is fairly constrained, and user interaction can be reduced to a selection process, then a more limited input device is acceptable. Concurrent tasking must also be considered; if the operator's visual attention is totally occupied by a task, then the use of gaze pointing for control is not appropriate.

Task-controller mapping

In addition to considering the adequacy of the candidate input device in general, the specific mapping of the input device to control functions must be addressed. It is unlikely that a single hands-free input device will be adequate for all control functions required for wearable computers. A specific input device will be elegantly appropriate for some control functions and clearly inappropriate for others. It is most likely that alternative controllers will be used in conjunction with conventional manual input devices and other alternative controllers. Thus, task-controller mapping must take into account how best to increase overall functionality by using multiple input devices. The following paragraphs describe some mapping alternatives. Wearable computers may incorporate more than one of these mapping techniques in the overall control system design.

Single input device mapped to single control function. With this mapping, the "optimal" input device is assigned to each control function. For example, one might use keyboard control for inputting lengthy text information and a hands-free controller for sequencing through display screens or options. The potential advantages of using optimized input devices needs to be weighed against controller cost and potential operator confusion concerning which device to use for each function.

Single input device mapped to multiple control functions. It is doubtful that any hands-free input device will be capable of performing *all* the control functions required for wearable computer operation. To the extent that this can be achieved, controller cost can be reduced. However, the designer needs to avoid the pitfall of compromising overall efficiency for the sake of using fewer devices. For instance, speech-based control currently offers the most capability for this type of mapping. However, use of a speech command to designate a position on a two-dimensional surface can be cumbersome.

Multiple input devices mapped to single control function. Just as operators with desktop computers can navigate with a variety of controllers (mouse movement, arrow keys, tab key), it is possible to implement wearable computer operations such that several input devices can be used for a single control action. This mapping approach provides the operator with increased flexibility: a) the operator may have individual preferences for specific input devices, b) a temporary task or environmental condition may deem one input device more efficient than another, and c) should one device malfunction, the control action can still be performed with another device. Once again, the designer needs to ensure that overall efficiency is not compromised. To map one, less optimal, input device to a control function may require procedures (e.g., a hierarchical menu approach) that are not optimal for another device (e.g., single speech command). This requires the operator to use different procedures, depending on the active input device, for the same control function. There is a limit, though, to operators' ability to remember or manage multiple procedural steps.

Mappings that truly integrate multiple input devices. In certain cases a combination of two or more input devices can perform a control function better than either one operating alone. One method is to map the input devices to subcomponents of the control action. For example, if hands-free function selection is required, eye gaze can be used to designate a desired control function and a purposeful facial muscle signal can serve as a consent response, commanding the system to act on the function last designated by the eye gaze. It would be difficult to use the individual input devices for this overall function. The use of both input devices capitalizes on the ability of eye gaze to rapidly designate position on a two-dimensional surface and a muscle signal to quickly send a command.

A second method integrates multiple input devices to increase the accuracy or reliability of a control action. For example, the use of lip movement data, together with acoustic signals recorded during voice commands, has been found to enhance the performance of a speech recognizer in a noisy environment. Similarly, a controller design might require a simultaneous eye blink and muscle signal to minimize the chance of spurious activations.

A third method uses one input device to improve the performance of a second device. For example, eye line-of-sight data might be used to enhance speech processing by restricting the vocabulary search to the most probable verbal commands associated with the current gaze point. Although little research has been done with integrated mappings, it seems clear that these types of designs will best capitalize on the capabilities of hands-free input devices.

Frequency of control input

Although wearable computers are essentially "on" all the time, the frequency of operator input can range from constant to sporadic and can vary with task demands. Controller selection should take into consideration the anticipated input frequency. For example, just as extended manual keyboard entry can cause carpal tunnel syndrome, frequent use of jaw clenches to activate EMG-based control can aggravate TMJ (tempero-mandibular joint) disorders.

Controller activation

Input devices for wearable computers need to be either constantly operational or capable of being engaged in minimal time. As outlined in previous examples, the use of a multimodal interface allows one modality to serve as an activation command or consent for another input modality. For instance, continuous modulation of icons in the EEG-based input device described earlier may be too distracting to the operator. A purposeful facial muscle signal could be used to start modulating the icons when the operator desires to make a control input and terminate the modulation when the control action is finished. Any distraction experienced during the brief activation time should be outweighed by the benefits of using a hands-free direct function selection approach.

Controller evaluation

Laboratory evaluation, together with field testing using target operators, are essential for the design of effective input devices and task-controller mappings. Objective measures of performance (e.g., time to activate the device, task completion time and accuracy, and performance on concurrent tasks) should be recorded, in addition to subjective ratings on the usability, satisfaction, and learning associated with the candidate controller. Workload measures can also indicate whether the input device facilitates computer operation and whether it diverts attention from the primary task.

Conclusions

The application of hands-free input devices for wearable computer applications is an endeavor in both science and art. Further hardware and software developments are required in the enabling technologies. Creative human-computer dialogue design, together with human factors research and field testing, is required to evolve a system design that exploits the combined potential of manual and hands-free controls. In examining potential applications of hands-free input devices, the opportunity for developing innovative human-computer dialogues should be seriously considered. Most existing dialogues are tailored for a graphical user interface and mouse controller. With hands-free control, novel dialogues are now possible, and will be required to achieve the maximum benefit.

With careful design, there are numerous advantages to using nonconventional controllers with wearable computers. Hands-free operation can be realized for many tasks. Given the variety of available input modalities, there will useful options for any special user population (e.g., those with severe physical limitations). Properly implemented, the human-machine interface will be more natural and require less training. In many cases, inputs can be made with more speed and accuracy. Given these advantages, hands-free input devices have great potential to assist operators in benefiting from the full capabilities afforded by wearable computer technology.

References

[1] Bass, L., Kasabach, C., Martin, R., Siewiorek, D., Smailagic, A., and Stivoric, J. The design of a wearable computer. *Proceedings of CHI '97* (Atlanta, GA, March 1997), ACM Press, 139-146.

[2] Boeing Wearable Computer Workshop breakout Session Summary. See http://www.cs.cmu.edu/afs/cs.cmu.edu/project/vuman/www/boeing/index.html.

[3] Borah, J. Helmet-Mounted Eye Tracking for Panoramic Display Systems - Volume I: Review of Current Eye Movement Measurement Technology (AAMRL-TR-89-019) (NTIS: AD-A273 101/6/XAB), 1989.

[4] Calhoun, G.L., Janson, W.P., and Arbak, C.J. Use of eye control to select switches. *Proceedings of the 1986 Human Factors Society Annual Meeting* (Dayton, OH, October 1986), Human Factors Society, 154-158.

[5] Farwell, L.A., and Donchin, E. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 70, (1984), 510-523.

[6] Fukumoto, M., and Tonomura, Y. Body coupled FingeRing: Wireless wearable keyboard, *Proceedings of CHI '97* (Atlanta, GA, March 1997), ACM Press, 147-154.

[7] "In the Blink of an Eye," The Human Element: Center for Human Performance in Complex Systems Newsletter, 2(1), (January 1997), 1-2.

[8] Jennings, D.L., and Ruck, D.W. Enhancing automatic speech recognition with an ultrasonic lip motion detector. *Proceedings of the IEEE International Conf. on Acoustics, Speech and Signal Processing* (Detroit, MI, 1995), 868-871.

[9] Lai, J., and Vergo, J. MedSpeak: Report creation with continuous speech recognition. *Proceedings of CHI '97* (Atlanta, GA, March 1997), ACM Press, 431-438.

[10] Matias, E., MacKenzie, I.S., and Buxton, W.W. A wearable computer for use in microgravity space and other non-desktop environments. *Proceedings of CHI '96* (Vancouver BC, Canada, April 1996), ACM Press, 69-70.

[11] McMillan, G.R., Eggleston, R.G., and Anderson, T.R. Nonconventional controls. In G. Salvendy (Ed), *Handbook of Human Factors and Ergonomics*, 2nd edition (1997), John Wiley and Sons, New York, 729-771.

[12] Nasman, V.T., Calhoun, G.L., and McMillan, G.R. Brain-actuated control and HMDs. In J. Melzer and K. Moffitt (Eds), *Head-Mounted Displays: Designing for the User* (1997), McGraw-Hill, New York, 285-310.

[13] Nelson, W.T., Hettinger, L.J., Cunningham, J.A., Roe, M.M., Lu, L.G., Haas, M.W., Dennis, L.B., Pick, H.L., Junker, A., and Berg, C.B. Brain-body-actuated control: Assessment of an alternative control technology for virtual environments. *Proceedings of the 1996 IMAGE Conference* (Scottsdale, Arizona, June 1996), 225-232.

[14] newAbilities UCS 1000. See http://server.seaside.org/9501.html.

[15] Salem, C., and Zhai, S. An isometric tongue pointing device. *Proceedings of CHI '97* (Atlanta, GA, March 1997), ACM Press, 538-539.

[16] Sturman, D.J., and Zeltzer, D. A survey of glove-based input. *IEEE Computer Graphics and Applications*, 23, (1994), 30-39.