Foundations and Trends® in
Human–Computer Interaction
Vol. 4, No. 4 (2010) 245–316

DOI: 10.1561/1100000023

# Emerging Input Technologies for Always-Available Mobile Interaction

By Dan Morris, T. Scott Saponas, and Desney Tan

## Contents

Foundations and Trends® in
Human–Computer Interaction
Vol. 4, No. 4 (2010) 245–316

DOI: 10.1561/1100000023

# Emerging Input Technologies for Always-Available Mobile Interaction

**Dan Morris, T. Scott Saponas, and Desney Tan**

*Computational User Experiences Group, Microsoft Research, USA*
{*dan, ssaponas, desney*}*@microsoft.com*

## Abstract

Miniaturizing our computers so we can carry them in our pockets has drastically changed the way we use technology. However, mobile computing is often peripheral to the act of operating in the real world, and the form factor of today's mobile devices limits their seamless integration into real-world tasks. Interacting with a mobile phone, for example, demands both visual and manual focus. We describe our goal of creating *always-available interaction*, which allows us to transition between mobile computing and real-world tasks as efficiently as we can shift our visual attention. We assert that this could have the same magnitude of impact that mobile computing had on enabling tasks that were not possible with traditional desktop computers.

In this review, we survey and characterize the properties of sensors and input systems that may enable this shift to always-available computing. Following this, we briefly explore emerging *output* technologies, both visual and non-visual. We close with a discussion of the challenges that span various technologies, such as ambiguity, sensor fusion, gesture design, and cognitive interference, as well as the opportunities for high-impact research those challenges offer.

# 1
# Introduction

With recent advances in mobile computing, we have miniaturized our computers so we can carry them in our pockets (or bags or clip them on our clothes) and have relatively convenient access to information and computation even when we are not sitting at our desks. This has drastically changed the way we use technology and has impacted our work and life in profound ways. However, contrary to computing being the primary and only task in desktop scenarios, computing in mobile scenarios is often peripheral to the act of operating in the real world. We believe that there remain opportunities for more tightly infusing computational access into our everyday tasks.

At present, the form factor of typical mobile devices limits their seamless integration into real-world tasks: interacting with a mobile phone, for example, demands both visual and manual focus. For example, researchers have shown that users could attend to mobile interaction bursts in chunks of about 4–6 seconds before having to refocus attentional resources on their real-world activity [97]. At this point, the dual task becomes cognitively taxing as users are constantly interrupted by having to move focus back and forth. Unfortunately, when Ashbrook et al. measured the overhead associated with mobile interactions, they found that just getting a phone out of the pocket or hip holster takes about 4 seconds, and initiating interaction with the device takes another second [5]. This suggests that the current status quo in mobile interaction will not allow us to integrate computing tightly with our everyday tasks.

In our work, we assert that augmenting users with always-available interaction capabilities could have impact on the same magnitude that mobile computing had on enabling tasks that were never before possible with traditional desktop computers. After all, who would have imagined mobile phones would make the previously onerous task of arranging to meet a group of friends for a movie a breeze? Who would have imagined when mobile data access became prevalent that we'd be able to price shop on-the-fly? Or resolve a bar debate on sports statistics with a quick Wikipedia search? Imagine what we could enable with seamless and even greater access to information and computing power.

We spend a majority of this review surveying the state of the art in novel input modalities that may allow us to transition between physically interacting with the mobile device and with the real world as efficiently as we can shift our visual attention back and forth between the two. We specifically assert that certain input technologies are more likely than others to play a role in this paradigm shift, and attempt to characterize the properties of sensors and input systems that render them promising for always-available computing. Although this article's focus is on input technologies, efficient micro-interaction will also require an approach to *output* that is less cognitively demanding than current mobile displays. We thus follow our input-technology survey with a brief exploration of emerging output technologies, both visual and non-visual. After surveying and characterizing these technologies, we close the review with discussion of challenges that span diverse technologies, such as systematically handling ambiguity, sensor fusion, gesture design and applicability, and cognitive interference associated with using them in the real world, as well as the opportunities for high-impact research those challenges offer.

# 2

# Always-Available Input Technologies

In this review, we aim to provide broad appreciation for historical input research, but to focus most of our effort on more recent technologies and techniques we feel to be relevant to attaining "always-available mobile micro-interactions." Our first goal, therefore, is to scope our survey and informally outline several requirements for always-available mobile input that enable micro-interactions:

(1) Always-available input may require a cognitive shift to the task for which the user demands input, but the input modality itself should not disrupt cognition. Just as I can be engaged in conversation and briefly pause to tie my shoe or say "hello" to a third party, an always-available input system should require only the amount of distraction that the underlying computing task introduces.
(2) Transitioning in and out of always-available input should be as rapid as transitioning our visual attention from one task to another. If an input system takes 10 seconds to access, it is not "always-available".
(3) Always-available input should be portable to any environment, within reason. While we do not argue that even the

optimal technologies will work underwater, for example, we do argue that technologies exist to provide always-available input in environments that are novel to the user, both indoors and outdoors.

(4) The notion of "always-available" includes scenarios when a user's hands are busy with other tasks: therefore, always-available input should be at least *compatible* with the use of our hands for non-computer-based tasks.

Always-available methods can range in the bandwidth of communication they support. While we believe that useful applications can be crafted around the full range, in this review, we slightly favor modalities that provide higher bandwidth (e.g., preferring techniques that use detailed finger gestures over whole-body gestures). But ultimately, the goal of any input technique is to capture the intent of the user and transform it into actions that the computer can perform. We thus organize subsequent subsections around sensors or input modalities we find most promising for capturing user intent in mobile scenarios.

It should also be noted that one may dichotomize the sensor space into sensors that are placed in the environment and ones that are placed on the human body. While it may be reasonable to assume that certain environmental sensors will become so prevalent as to pervade all our computing environments, we assert that the list of interesting computing environments is constantly growing, and that there exist significant mass deployment challenges to do this. In this survey, we favor describing the emergence of technologies that are carried or worn on the body and that are truly mobile, and leave survey of environmental sensors as well as projections of the eventual integration for a separate review. In reading through the survey, we urge the reader to consider the infrastructure and critical mass required (or not) for deploying some of these mobile sensors, as well as to imagine the integration of these technologies with more traditional, infrastructure-dependent technologies.

To keep logical order, we begin with technologies that are already in use today, such as inertial sensors and touch input, and proceed with technologies that we see as increasingly forward-looking.

## 2.1 Inertial Motion Sensing

We begin by looking at sensors that measure their own movement, which — when held in a user's hand or worn on a user's body — allows computers to measure movement related to users' gestures or physical activities.

As its name suggests, an accelerometer is any sensor that measures its own acceleration. The most common accelerometer design used for computer input consists of a damped mass on a spring in a rigid housing. When an external force accelerates the whole system, the mass is displaced relative to its housing, and this displacement — which is proportional to acceleration — is measured. This method is relatively easy to fabricate at small scales (e.g., using microelectro-mechanical systems, or MEMS), and is simple, reliable, and inexpensive. Most micromechanical accelerometers are designed to be sensitive only to a single direction in one plane. By integrating multiple devices perpendicularly, two- or three-axis accelerometers can be made.

Gyroscopes are sensors that measure changes in their own *orientation*, and may be constructed using any of several operating principles. Rotating-disk gyroscopes, for example, are built by placing a spinning disk inside a non-spinning housing. When the entire gyroscope is rotated, the disk's inertia tends to keep it spinning in its original plane of rotation, creating a torque or displacement between the spinning and non-spinning parts of the gyroscope. This torque is directly related to the applied rotation. Vibrating planes have a similar tendency to resist rotation, a principle that allows gyroscopes to be built based on vibrating piezoelectric materials. The latter approach is frequently applied in modern MEMS gyroscopes, where the vibrating element is embedded in a silicon die along with the electronics required to measure its displacement.

The combination of these technologies has been used for sensing ranging from braking systems in cars to monitoring commercial machinery to medical applications to navigation and guidance systems. They are also starting to see mass deployment in many modern mobile phones. While it is outside the scope of this review to survey all applications of these sensors, we focus this subsection on recent uses that include hand-held devices, and perhaps more interestingly, body-worn devices.

Perhaps the most popular recent commercial success to utilize accelerometers for computer input is the controller that comes with Nintendo's Wii game console.[1] This controller utilizes a combination of optical sensors and accelerometers to provide motion sensing capability, the basis for interaction with this device. This concept had been previously explored in academic circles: for example, Wilson and Shafer describe a hardware device called the XWand that used a two-axis accelerometer, a three-axis magnetometer, and a single-axis gyroscope to sense gestures and pointing direction [148] (Figure 2.1). The use of inertial sensing for entertainment has achieved even further success with the incorporation of accelerometers and gyroscopes into mobile phones, which now leverage these sensors for gaming, music synthesis, pedometry, and a variety of other applications.

This popularization of accelerometer-based entertainment devices and applications has inspired academic work as well, exploring various aspects of tracking user motion with accelerometers. For example, Rehm et al. use the Wii Remote for exploring cultural influences on

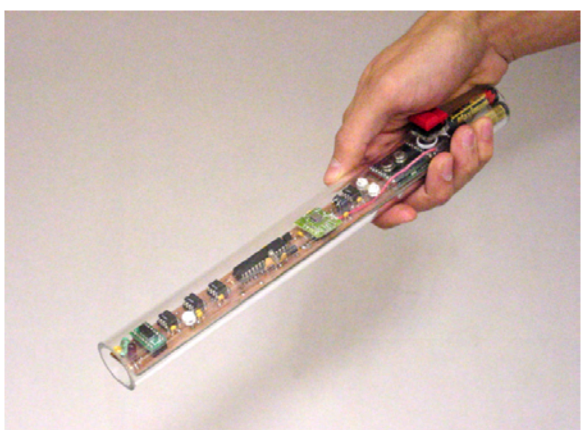

Fig. 2.1 Wilson'sand Shafer's XWand [148] combined accelerometers with a gyroscope and a magnetometer. Image © ACM 2003.

[1] Nintendo Co., Ltd., http://www.nintendo.com/wii.

gestural execution [110]. Furthermore, many researchers have worked on the core problems associated with taking the raw sensor data and performing gesture tracking and inference. Pylvanainen describes in great detail using hidden Markov Models to infer accelerometer data [107], for example deriving optimal ways to normalize and rotate an accelerometer data vector to get it aligned to the universal frame, i.e., aligning the y-axis with gravity.

Recently, researchers have begun to explore attaching these sensors to the body to track motion in order to control various applications. We believe that this is a promising path to providing always-available input. Using accelerometers on the wrist and arm, Cho et al. decode gestures for emulating devices like TV remote controls [20] (Figure 2.2). They focus on low-power processing and take an interesting heuristic approach in which they manually classify which planes (XY, YZ, XZ) are traversed by each of their gestures. They report attaining a 73% recognition rate for 12 gestures, and find that mounting sensors on the wrist works better than on other parts of the arm. Other researchers have looked at sensing more minute finger gestures with minimal instrumentation. In 1994, Fukumoto and Suenaga leveraged single-axis accelerometers on each finger to detect when the fingers strike


Fig. 2.2 Cho et al. [20] apply lower-power, plane-crossing-based techniques to wrist- and arm-mounted accelerometers. Image courtesy of authors.

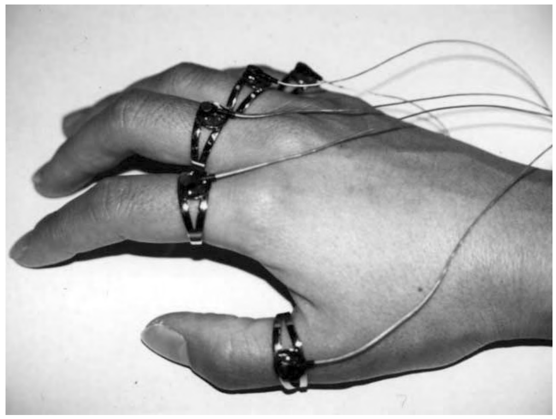

Fig. 2.3 Fukumoto and Suenaga [32] placed accelerometers on the fingers to detect gestures and contact with surfaces. © ACM 1994.

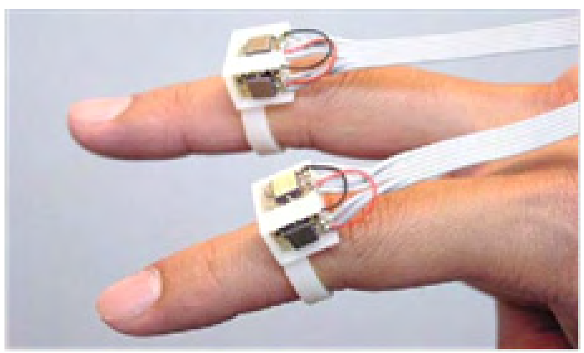

Fig. 2.4 Lam et al. [72] placed accelerometers on rings, wired to a radio transmitter (not shown) worn on the wrist. © IEEE 2003.

a surface [32] (Figure 2.3). They use this to provide text input through a chorded input mechanism. Likewise, Lam et al. [72] (Figure 2.4) use rings fitted with accelerometers. The rings are worn on four fingers, with cables running to a wireless transmitter worn on the wrist.

Because inertial sensing has become almost ubiquitous in mobile phones, applications have also begun to emerge in which a phone's embedded accelerometers and gyros are used for always-available input. Though we have specifically scoped our definition of "always-available" to exclude scenarios where a user has to reach into his pocket to access a device, Hudson et al. [53] leverage a phone's built-in accelerometer even while it is still in the user's pocket, to sense what they label "Whack gestures". Using their technique, a user can slap the device that is still residing in a pocket or backpack to communicate a small amount of information, essentially enabling low-bandwidth, but always-available, interaction. For example, a user might "whack" a phone that is still in his or her pocket to silence the phone's ringer.

## 2.2 Touch Sensing

While inertial sensing already plays an important role in mobile input, it may lack the precision for high-bandwidth tasks like text entry and object selection. In this section, we will discuss technologies that leverage our precise control over our fingers for mobile input, using both mechanical and electrical sensing.

A conventional technology that has been explored for mobile input is the button-based keyboard: various conceptualizations of the keyboard have enabled lightweight, one-handed, mobile use through chording and sequencing (e.g., the Twiddler,[2] studied in Ref. [78]) (Figure 2.5). This approach has achieved significant commercial success for mobile input, particularly since the advent of T9[3] and similar predictive input schemes. However, we do not believe that input techniques requiring a device held in a user's hand can be truly "always-available": the time required to access such a device (whether a phone or a standalone keyboard), and the incompatibility with any tasks occupying the user's hands or otherwise prohibiting manual interaction, separate keyboard-based input from truly always-available input.

Moving beyond mechanical buttons and actuators, which necessarily separate the input and output media used for interaction, various sensing technologies allow users to interact directly on the surface used for information display. While traditional mechanical buttons also of course require the user to "touch" the input device, "touch input" has come to refer to these technologies that allow co-located display-based output and touch-based input. Chang et al. overview of some of the sensing

[2] Handykey Corporation, http://www.handykey.com.
[3] Nuance Communications, Inc., http://www.t9.com.


Fig. 2.5 The Twiddler hand-held keyboard allows one-handed, mobile text entry using physical buttons. Image from Lyons et al. [78], © IEEE 2004.

mechanisms underlying touch-input systems [18]. These include resistive and capacitive sensors, surface acoustic wave transmission, and infrared or color cameras. Some of these mechanisms can be equally applied to either the finger or a stylus, while others apply exclusively to one or the other.

Bill Buxton provides historical perspective on touch-sensitive devices in [16], capturing the evolution of touch input from mechanical transducers to what we know today as touch-sensitive surfaces (e.g., Microsoft Surface 2.0[4]). He additionally postulates several shortcomings of touchscreen technologies, namely: (a) the sole reliance on visual feedback to operate the interface means that if you are blind or otherwise cannot focus visual attention to the display, you cannot use this interaction style; (b) even when you can dedicate visual attention,

[4] Microsoft Corp., http://www.microsoft.com/surface/.

many of the displays attached to these devices do not work well in extreme lighting conditions, such as under direct sunlight; (c) virtually all handhelds relying on touchscreens require both hands to operate; and (d) finger interaction is generally much less precise than interaction with a physical stylus.

Hence, we make the same argument regarding the increasingly ubiquitous mobile touchscreen, most commonly constructed using capacitive sensing. While touch-based mobile devices are extremely portable, they are still devices that a user must retrieve from storage prior to interaction (a time-consuming operation relative to the fluidity of our cognitive and visual attention), and they are still obstructions to everyday tasks that require our hands. I can send a text message from my mobile phone or I can carry my grocery bags, ride my bike, walk my dog, hold my child, etc. But using modern touch-based devices I cannot, so to speak, "have my hands and use them too."

Interestingly, Saponas et al. present a technique called PocketTouch that allows a user to interact with their capacitive touchscreen through fabric, that is, without ever taking the device out of their pocket or bag [124]. Various other researchers have been pushing instead on the boundary of touch screens that do not require storage and that stay out all the time to achieve the always-available vision. The *nanoTouch* project explored techniques for interacting on the back of devices with extremely small screens (~2.4 inch) [11] (Figure 2.6). This eliminated occlusion of the screen by the fingers, and opened an area of study motivated by creating devices that never had to be put away. Extending observations from this work, Holz and Baudisch describe *RidgePad*, a touch sensing technique that records the user's fingerprint on the screen, in addition to basic positional data. This not only provides user identification, but also uses the inferred 3D posture of the finger to improve tracking [49] (Figure 2.7). Asserting that the wristwatch is a device that is quick to access for micro-interactions, Ashbrook et al. explore interaction techniques based on a circular touchscreen wristwatch [6]. They consider three types of inter-target movements for variously sized buttons placed around the rim, and derive a mathematical model for error rate given a movement type and angular and radial button widths.

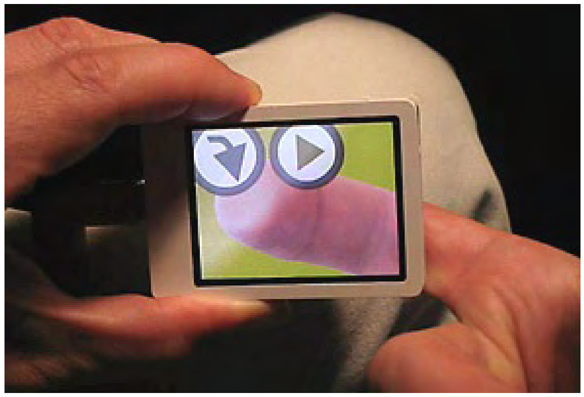

Fig. 2.6 Baudisch et al. explore interactions on the back of very small devices in their nanoTouch project [11]. © ACM 2009.

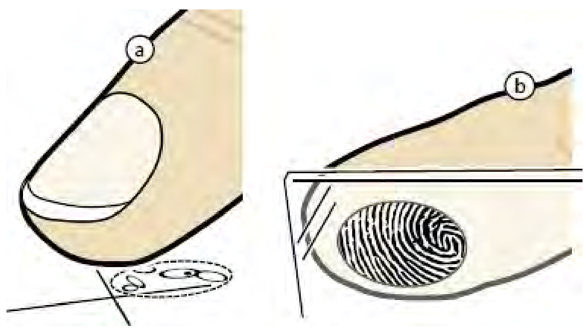


Fig. 2.7 Baudisch et al.'s *RidgePad* [49] project uses fingerprint patterns to improved finger tracking for touch input. © ACM 2010.

More generally, Ni and Baudisch survey candidate techniques for gesture-based interactions with "disappearing mobile devices" [91]. They report on results of two studies investigating affordances of these devices, focusing on marking and text entry using a gesture alphabet. Similarly, Gustafson et al. describe Imaginary Interfaces which is a concept they used to explore the extent to which users could spatially interact with screen-less devices and interfaces that existed only in the imagination [40] (Figure 2.8). They find that short-term memory could at least partially replace conventional visual feedback, and that users could create simple drawings, annotate existing drawings, and point at precise locations described in imaginary space.

Fig. 2.8 Gustafson et al. [41] explore users' execution of gestures for "Imaginary Interfaces." © ACM 2010.

Beyond mobile input devices, touch-sensitive surfaces integrated into walls and furniture have recently received tremendous academic and commercial attention (e.g., [41], Perceptive Pixel's Multi-touch Display and Multi-touch Wall,[5] Microsoft Surface[6]). While this approach offers increasingly natural input, it depends on significant environmental modification, and — even more than camera-based sensing — constrains the location and behavior of a user even *within* an instrumented environment. Therefore, we similarly expect that environmentally instrumented surfaces are unlikely to be central to the emergence of always-available micro-interactions.

## 2.3 Computer Vision

While touch sensing offers high precision for two-dimensional interactions, it limits the interaction space to a physical surface. This constrains both the environments in which touch will be practical (a user needs to be able to approach and manipulate the input device) and the vocabulary of possible gestures the modality can support. In this section, we explore techniques that use computer vision to extend user

[5] Perceptive Pixel, Inc. http://www.perceptivepixel.com.
[6] Microsoft Corp., http://www.microsoft.com/surface.

input to three dimensions and relax the requirement that input requires mechanical contact with a sensor.

Computer vision — roughly defined as analyzing patterns collected from an array of light sensors — has received extensive attention from computer scientists for applications ranging from medical image analysis to robot navigation. And perhaps no technology has received more attention as a means to hands-free interaction between humans and computers. Note that we say "hands-free" here, rather than "always-available". In this section, we will explore this dichotomy, and discuss a variety of ways in which computer vision can be applied to HCI, and ultimately to mobile interaction.

### 2.3.1 Environmentally-Situated Cameras

First and foremost, the HCI and computer vision communities have extensively explored the use of environmentally situated cameras for analyzing gestures, particularly hand gestures (e.g., [13, 69, 81, 104, 135]). This approach has received some commercial success as well, particularly through incorporation into gaming consoles, most notably the PlayStation Eye,[7] which leverages a color camera for coarse gesture interpretation, and the Nintendo Wii,[8] which leverages a handheld infrared (IR) camera and an environment-mounted IR emitter to localize the handheld device relative to the emitter. Wachs et al. survey sensing technologies for hand gesture recognition based on environmental cameras, and discuss emerging applications for recognizing hand gestures [99].

More recently, the use of "cameras" to interpret user input has been broadened to include vision-based 3D sensors (e.g., [150, 149]). Microsoft's Kinect[9] represents perhaps the first application of this approach to consumer scenarios, leveraging a depth-sensing camera for gesture interpretation.

While the use of environmentally situated cameras is promising for scenarios where a camera is available, and while the decreasing cost

[7] Sony Computer Entertainment, http://us.playstation.com/ps3/accessories/playstation-eye-camera-ps3.html.

[8] Nintendo Co., Ltd., http://www.nintendo.com/wii.

[9] Microsoft Corp., http://www.xbox.com/en-US/kinect.

of charge-coupled devices (CCDs) and other light sensors allows cameras to be deployed in more and more environments, we argue that the environment will never be sufficiently instrumented to build "always-available" interactions around environmentally-situated cameras. We thus devote the remainder of this section to applications of vision-based interaction that have perhaps received less attention than environmentally situated cameras, but may offer a more feasible path to always-available interaction.

### 2.3.2 Gaze Tracking

The majority of the technologies we discuss in this survey leverage our hands for input. However, significant industrial and academic attention has also been paid to using our eyes for input, through various forms of gaze tracking (e.g., [42, 28, 29, 86, 90]). Mobile approaches to eye tracking are even becoming plausible, as an increasing set of mobile devices — and potentially even glasses or contact lenses — incorporate sensors capable of following a user's gaze. This holds significant promise not only for motor-impaired users, but also for collecting implicit information about a user's attention. But it is precisely that tight implicit link to attention that we argue prohibits the use of gaze tracking in always-available input systems. However, since it is hard for a human to decouple their eyes from their attention, it remains difficult to harness gaze as a conscious input stream.

### 2.3.3 On-Body Cameras

Although environmentally situated cameras offer great potential for hand gesture recognition, we argue above that this approach will not generalize to always-available, mobile interaction. However, an alternative approach — *mounting cameras on a user's body* — may leverage the potential of computer vision in a mobile input system. The *Sixth-Sense* project [86] (Figure 2.9), for example, envisions a color camera worn in a hat or pendant that looks down on a user's hands, sensing and interpreting hand gestures in any environment. This work proposes the incorporation of a head- or pendant-mounted projector that would allow not only in-air gestures, but also interaction with a projected user


Fig. 2.9 The SixthSense project [86] envisions a color camera worn on a pendant that uses computer vision to recognize hand gestures. Courtesy Pranav Mistry.

interface. A similar form factor is used by Starner et al.'s *Gesture Pendant* [133] (Figure 2.10), which employs a neck-worn infrared camera and emitter for illuminating and sensing the hand.

Ahmad and Musilek [1] (Figure 2.11) explore a different form factor, mounting a camera on the palm side of a user's wrist, pointing toward the hand. The camera monitors the fingertips and can classify finger movements in two dimensions, offering a vision-based approach to capturing finger gestures that is perhaps applicable to always-available interaction. This system also demonstrates the use of arm movement for continuous control (e.g., cursor movement) using the same sensor configuration. A user can move his/her fingers out of the camera's field of view and switch the system into a pointing mode, in which the camera looks at the scene in front of the user and maps optic flow (an estimation of the overall movement of the scene, which in this case corresponds to arm movement, since the camera is attached to the user's arm) to cursor position.

Fig. 2.10 Starner et al.'s Gesture Pendant uses a neck-worn IR camera and emitter for illuminating and sensing the hand [133]. © IEEE 2000.

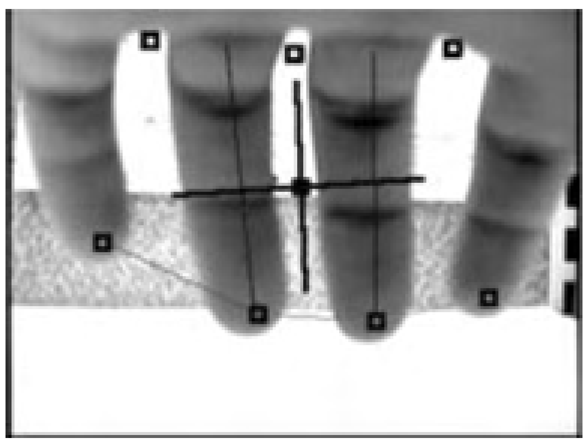

Fig. 2.11 Ahmad and Musilek [1] place a camera underneath the wrist, looking out toward the hand, to interpret finger and arm gestures. © IEEE 2006.

## 2.4 Mouth-Based Interfaces: Speech and Beyond

The majority of the technologies we have discussed so far attempt to leverage our manual dexterity for computer input, building on the legacy of hand-and-finger-based input devices (particularly the mouse and keyboard) but addressing mobile scenarios where traditional devices are impractical. However, all of these approaches share a common drawback: many real-world tasks require one or both hands, which sets a boundary on the scenarios where hand-based interaction with a computer will be appropriate. We can, however, decouple another important motor sub-system — speech — from many of our everyday tasks that require our hands. Controlling a computer via speech input does not require our hands and is extraordinarily portable, and — in many cases — does not interfere with "real-world" tasks. I *can* communicate with my computer using speech while I carry the groceries, ride my bike, drive a car, etc. In fact, perhaps the most successful commercial application of voice recognition in mobile environments is the use of voice input to dial and manipulate a phone while operating a vehicle. It would thus seem that speech is optimally poised to enable the mobile micro-interactions that we propose will lead to always-available computing.

However, other drawbacks of speech input render it unsuitable as a modality for micro-interaction. Most notably, *conversation* is perhaps the most precious "real-world" activity with which we would like our computing *not* to interfere, a requirement that speech input will almost certainly be unable to meet. Less obvious, perhaps, is the fact that the verbal nature of the human stream of consciousness results in a high level of interference between the use of speech and almost any cognitive task (Shneiderman [128] discusses the cognitive limitations associated with speech recognition interfaces). Furthermore, significant technical limitations call into question the ultimate performance of speech interfaces in real-world environments, and the strong association between social interactions and speech has raised further criticism of the role of speech in UIs. Starner [134] breaks down some of these social and technical limitations in more detail.

We therefore label speech recognition as perhaps the most controversial of the technologies we discuss in this review, in terms of its long-term role in mobile interaction. In this section, we will highlight recent approaches to bringing speech recognition to mobile interfaces, but further research is required to determine whether speech-based interfaces will be viable for mobile interactions in arbitrary environments.

### 2.4.1 Speech Input on Mobile Devices

Speech input does not represent an emerging sensor or modality per se, so a complete discussion of speech recognition in mobile environments is beyond the scope of this review. However, speech technology continues to evolve and has so much to offer for always-available interaction that we will overview some of the main challenges and trends in this space.

First and foremost, the traditional problem of transcribing even highly-controlled speech patterns comes with a host of challenges that are unique to the mobile space (Cohen [23] discusses some of the major issues faced by commercial efforts in mobile speech recognition). Low-level acoustic modeling becomes fundamentally more difficult than in the desktop or telephony spaces, due to unpredictable and often-noisy environments, and unpredictable and often less-than-optimal placement of the microphone relative to the user. This is of course magnified in the scenarios we focus on in this review: speech recognition is challenging enough even when a user can be expected to hold a microphone close to his or her mouth. In many "always-available" scenarios, we might not be able to make this assumption, greatly exacerbating signal-to-noise problems. Furthermore, mobile scenarios often come with the challenge of restricted computational resources or an increased reliance on network connectivity. Consequently, recent algorithmic research focuses not only on traditional speech recognition problems, but also on adaptations that are specifically necessary for the mobile scenario, for example the reduction of sporadic noise [50], the fusion of multiple recognition algorithms to increase robustness [89], and low-computational-cost speaker adaption [73].

In addition to these low-level acoustic issues, using speech for always-available mobile interactions poses some higher-level challenges. For example, one of our criteria for always-available interaction is minimal cognitive overhead associated with transitioning in and out of interaction. Fulfilling this requirement for speech input requires not only accurate recognition, but also accurate recognition of natural speech patterns that do not require the user to concentrate on producing recognizable speech. As a result, research on handling natural expressions of uncertainty (such as "something like" or "I don't know") [102] and research on robust recognition in the presence of vocalized hesitation (such as "ummm" or "errr") [36] will be critical for the success of even the simplest always-available mobile speech interfaces.

Furthermore, an always-available, speech-based interface needs to be listening constantly to its environment, which comes with a particularly difficult segmentation problem: the system has to differentiate ambient conversation from commands or dictation intended for the interface. This is especially difficult for dictation or text entry interfaces that need to handle interruptions in dictation that are demarcated only by affective qualities in the user's voice (e.g., a change in my tone as I order my coffee in the middle of dictating an email). But even for "command and control" interfaces in which the system need only interpret a finite set of commands, those commands may be embedded in natural speech as well (e.g., I might just be talking about "checking my email", without wanting to actually check my email) or at the very least may be acoustically similar to conversational speech that should not be interpreted by the system. Consequently, proper handling of continuous audio for speech recognition is another problem that will be critical to always-available speech interfaces. Paek et al., for example, explore probabilistic models for continuous listening [101], while Lunsford et al. [76, 77] explore the behavioral and acoustic cues that can help distinguish system-directed from conversational speech.

Finally, always-available speech interfaces will need to compensate for the fact that an error-prone or slow system breaks our "low cognitive overhead for transitions in and out of interaction" requirement just as surely as requiring unnatural speech. Therefore, research into graceful handling of errors and ambiguity at the UI level, and UI paradigms for rapid transition in and out of speech interfaces, is just as important as the aforementioned research into improving low-level recognition. For example Paek et al. [100] explore statistical models that predict a user's likely actions in a speech-based UI, which both improves recognition and reduces the net interaction time for frequent interactions, and Paek et al. [103] propose a mechanism for graceful fallback to another modality (in this case touch) when speech recognition errors or uncertainties occur. Goto et al. [35] address the "continuous listening" problem through a novel UI paradigm: allowing the user to indicate system-directed speech by varying vocal pitch, and further extend this work to incorporate other "meta-speech" cues, particularly the inclusion of non-vocalized utterances to delineate system-directed speech [37, 38].

### 2.4.2 Non-Speech Voice Input

Though communicating with computers through speech is intuitive and high-bandwidth, many criticisms of speech as an interface mechanism stem from its social intrusiveness. In particular, speech-based interaction with a computer during a conversation is extremely unlikely, even in an era where typing and interacting with mobile devices during conversation has become commonplace. Given that this limitation may be deeply embedded in human social behavior, it is worth exploring approaches that leverage the benefits of speech input while bypassing this limitation.

In particular, recent work has shown that it is possible to detect speech-like movement of the mouth, face, and throat even when no sound — or sound that is inaudibly quiet — is produced. Denby et al. [27] provide an excellent overview of a variety of technologies that show promise in this area. Perhaps the most well-developed of these is surface electromyography (sEMG) applied to the face and throat, which attempts to recognize the patterns of muscle activation required to control speech production (Figure 2.12). Jorgensen and Dusan [61] explore sEMG-based subvocal speech detection, and Jorgensen and Binsted [60] even provide a demonstration of this approach used to drive a Web browser. Promising but less-developed approaches include the use of ultrasound images of the tongue and lips to recognize movements that indicate speech patterns [54], extremely sensitive microphones that detect whisper-level speech but reject environmental sound (non-audible murmur microphones) [140], and even the use of implantable brain–computer interfaces for monitoring the areas of the brain that are associated with low-level speech production [14]. While this invasive approach is not feasible for general-purpose applications, it does provide a window onto the low-level signals that control speech production and may inform the development of more practical approaches, such as Jorgensen et al.'s work in sEMG-based speech recognition.

Fig. 2.12 sEMG applied to the throat can recognize the patterns of muscle activation required to control speech production. Image courtesy NASA Ames Research Center, Dominic Hart.

These approaches all attempt to use a modality other than sound to detect the processes normally associated with speech. An alternative approach uses sounds produced by the mouth that are *not* normally considered speech to drive computer interfaces. While this approach does not necessarily address the social compatibility problems that speech-based interfaces pose in mobile scenarios, it does overcome another limitation of speech-based control: speech is an excellent mechanism for controlling discrete values (such as text streams), but does not offer a natural mode of control for continuous parameters. Recent lines of work attempt to overcome this limitation by recognizing continuous voice parameters that a user can easily control. For example, Harada et al. [43] map the volume of a user's voice to continuous parameters in a drawing application (e.g., brush size, opacity), a multimodal approach to harnessing non-spoken voice parameters. Igarashi and Hughes [57] harness both the volume and the pitch of a user's voice during held vowels that are embedded within a speech control stream. For example, this work allows a user to control a TV's volume by saying "volume up, ahhhh", where the volume continues to increase as long as the user says "ahhhh". This hybrid approach — part speech, part non-text voice — offers an interesting approach to overcoming at least one limitation of speech recognition: its inherently discrete nature.

### 2.4.3 Tongue Input

While not directly related to speech recognition, another approach that turns the mouth into an input device involves sensing tongue movement. While this approach does not offer the bandwidth of speech recognition, it does potentially offer subtlety and silence, and may be a valuable input modality for low-bandwidth discrete input. Challenges arise, of course, in non-intrusively instrumenting a user's mouth. This problem has not been definitively solved yet, but a variety of sensing approaches have been explored. Peng et al. [105] explore perhaps the most straightforward approach: mounting a series of buttons in a user's mouth on a retainer-like apparatus; they present a wireless system capable of sensing and transmitting activation events on five membrane-covered switches. Huo et al. present the "Tongue Drive" system [56], which uses a magnet secured to the tongue (by adhesive, piercing, or clip) and a head-mounted magnetic sensing system to monitor tongue movements. They demonstrate over 95% accuracy for six discrete gestures. Strujik employs a related approach with less external visibility, using a tongue-mounted magnet and retainer-mounted inductor coils to sense tongue movement [137]. Finally, Saponas et al. [115] (Figure 2.13) use retainer-mounted infrared sensor/emitter pairs to classify four discrete tongue gestures with >90% accuracy.

Most of these efforts to sense tongue movement are targeted toward accessibility applications, for example [55] explore the use of the *Tongue*

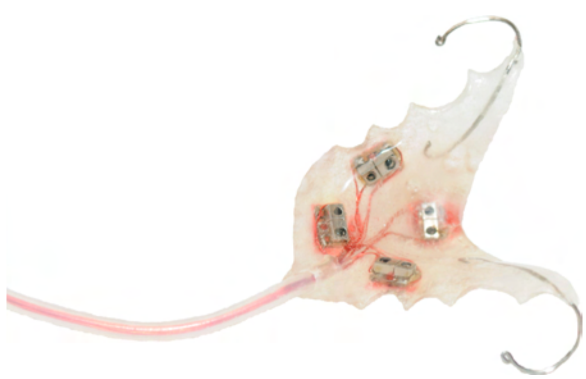

Fig. 2.13 Saponas et al. [125] use retainer-mounted infrared sensor and emitter pairs to classify tongue gestures. © ACM 2009.

*Drive* system as a control scheme for patients with spinal cord injuries. However, with sufficient miniaturization of the sensor apparatus, this approach may complement other approaches to always-available input by providing a covert, hands-free, discrete input stream.

## 2.5 Brain–Computer Interfaces

We asserted earlier in this review that "the goal of any input technique is to capture the intent of the user and transform it into actions that the computer can perform". What better way to capture a user's intent than to measure it directly, by capturing the electrical potentials that constitute a "thought" within the human brain? This is the broad goal of "brain–computer interfaces" (BCIs), and in this section we will discuss several sensing modalities used for sensing brain activity and their appropriateness for always-available, mobile micro-interactions.

Before discussing individual sensing technologies for BCIs, we will first summarize the theme of this section, by applying the criteria that we laid out above for always-available mobile input to brain–computer interfaces. Eventually, brain–computer interfaces may be fantastically appropriate for mobile interfaces: they provide by construct faster transitions in and out of communication with a computer than any other modality (requirement "2"), and they are by construct hands-free and compatible with a huge variety of physical tasks (requirement "4"). However, we believe that BCI technology may be several decades away from even letting us assess the practicality of brain–computer interfaces for everyday mobile interactions, and even further from deploying such interfaces. Sensors with sufficiently high bandwidth for most interface needs are prohibitively invasive, and non-invasive sensors have inadequate bandwidth and/or are prohibitively non-portable and expensive for real-world use. We do highlight that none of this precludes the applicability of these technologies to accessibility scenarios, where both portability and invasiveness need to be assessed against different criteria.

### 2.5.1 Implantable BCIs

Brain–computer interfaces may ultimately demonstrate the best performance when sensors are placed closest to the neurons (nerve cells)

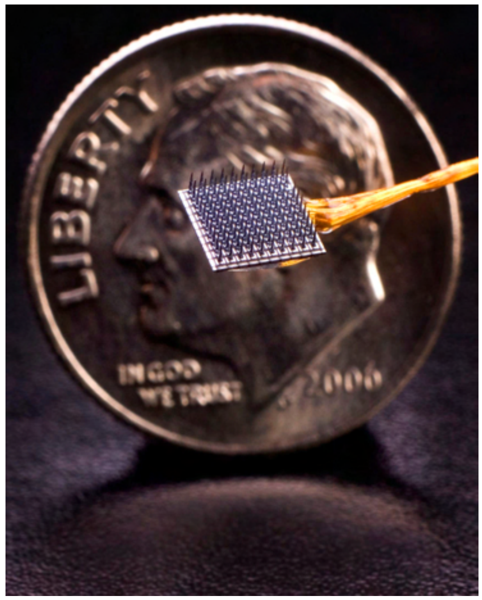


Fig. 2.14 Intracortical electrode arrays are surgically placed on the surface of the brain, and can record the electrical activity of several hundred brain cells. Image courtesy John Donoghue and Matthew McKee/BrainGate Collaboration.

that carry the brain's electrical signals. *Intracortical electrodes*, in particular, allow direct recording of the activity in a subset of the brain's neurons (Figure 2.14) through very small electrodes driven directly into brain tissue, so their tips are adjacent to individual brain cells. This approach offers relatively high bandwidth; in fact, implanted electrodes have allowed monkeys to directly control a three-dimensional cursor [152], and early results show that recording systems implanted in the brains of motor-impaired humans may also offer direct control of computer systems (e.g., on-screen cursors) [26, 68]. However, bioengineering challenges remain before these systems will be practical for long-term use [93, 113], and — more critically — this level of invasiveness is prohibitive for typical mobile interactions for the foreseeable future.

### 2.5.2 Electrocorticography (ECoG)

An intermediate level of invasiveness has attracted both research and clinical attention recently: *electrocorticography* (ECoG) uses electrodes

Fig. 2.15 Electrocorticography uses electrodes placed inside the skull — but not in brain tissue — to record activity from a large area of the brain. Image courtesy Eric Leuthardt.

placed on the surface of the brain (inside the skull) to record electrical activity with slightly lower temporal and spatial precision than intracortical electrodes (Figure 2.15), but with significantly less risk to patients. Although the idea of having anything placed on the surface of the brain may seem daunting and invasive to typical consumers, ECoG is considered only semi-invasive — and comparably quite safe — by neurosurgical standards, and has already achieved widespread clinical use for a variety applications. And as with intracortical electrodes, early results show that ECoG may have sufficient bandwidth for direct control of computer input signals [71, 106, 120], perhaps far more than electroencephalography (EEG), discussed below. However, despite being considered quite safe for clinical applications where the benefits far outweigh the risks, this approach is still prohibitively invasive for typical mobile input.

### 2.5.3 Electroencephalography (EEG)

In contrast to these more invasive, surgically-inserted sensors, electroencephalography (EEG) uses electrode plates on the surface of the scalp

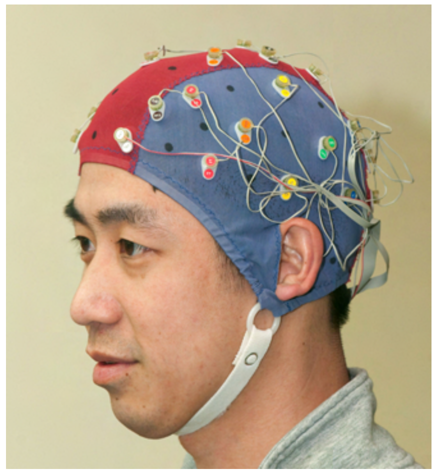

Fig. 2.16 Electroencephalography is non-invasive and relatively inexpensive, but does not provide high-resolution information about brain activity.

to record electrical activity from the brain (Figure 2.16). The benefit of this approach, relative to more invasive ECoG or implantable systems, is a huge reduction in invasiveness: EEG is safe and painless. Furthermore, EEG can potentially be quite inexpensive: although high-quality amplifiers are required to process the extremely weak signals measured on the scalp, such amplifiers are falling in cost thanks to their use in other applications, so an EEG system could potentially be made available at consumer price points [74]. And while EEG does require a grid of electrodes placed on the head, a complete EEG system is relatively portable (compared to fMRI or similar imaging technologies, discussed below). Most importantly, early evidence shows that EEG signals can be used to decode some degree of user intent, and in some cases EEG may offer sufficient bandwidth for direct control of a computer, particularly for accessibility scenarios [33, 82, 83, 121, 122]. EEG may even have potential to allow the control of three-dimensional continuous output signal [12].

All of this potential comes at a cost, though: a tremendous amount of detail is lost as electrical signals propagate through the skull and underlying tissue, leading to much lower spatial and temporal precision than ECoG or implant-based recordings. Consequently, EEG has very limited bandwidth and may ultimately be restricted to implicit sensing. Furthermore, although explicit control is *possible*, it requires intense focus at present and has not been shown to be feasible for a large slice of users [31]. In other words, using EEG for computer interfaces currently requires too much cognitive attention to be useful. Therefore, EEG may ultimately be most appropriate for implicit sensing, particularly in research environments, where it has already been harnessed for several implicit input paradigms: human-aided computer vision [64], cognitive load assessment [39], and task classification [74].

### 2.5.4 Functional Near-Infrared Spectroscopy (fNIRS)

*Functional near-infrared spectroscopy* (*fNIRS*) measures the reflectance of infrared light directed into the skull, which has been shown to vary with the underlying brain activity as a consequence of changes in blood flow patterns (Figure 2.17). fNIRS shares many properties with EEG: it is non-invasive, relatively cheap, and relatively portable, but lacks the spatial or temporal bandwidth required for direct control, and thus is likely unsuitable for mobile input. However, fNIRS — like EEG — holds tremendous potential as an implicit measurement tool for human–computer interaction (HCI) [34, 46, 92, 130, 119].

Fig. 2.17 Functional near-infrared spectroscopy is non-invasive and inexpensive, but probably lacks the bandwidth for direct-control brain–computer interfaces. © ACM 2009.

Fig. 2.18 Magnetoencephalography offers better temporal precision than EEG or fNIRS, but is non-portable and expensive, and probably still lacks the bandwidth for direct-control brain–computer interfaces. Courtesy National Institute of Mental Health, National Institutes of Health, Department of Health and Human Services.

### 2.5.5 Magnetoencephalography (MEG)

*Magnetoencephalography* (MEG) leverages the *magnetic* field created by the brain's electrical activity to assess brain activity (Figure 2.18). MEG offers better temporal precision than EEG or fNIRS, and is also non-invasive, but still likely lacks the spatial precision and overall bandwidth required for direct control applications. Perhaps more importantly, MEG equipment is extremely large and expensive (in fact requiring a magnetically-shielded room), with no clear path to reduction in size or cost, so it likely remains a research technology for the foreseeable future, and is an unlikely candidate for always-available input. With that said, the underlying data stream indeed contains information about motor intent [84, 145, 146] that may complement other technologies in motor control research.

### 2.5.6 Functional Magnetic Resonance Imaging (fMRI)

Finally, *functional magnetic resonance imaging* (*fMRI*) has received some attention in recent BCI research [109, 112] (Figure 2.19). fMRI leverages the fact that changes in neural activity in the brain result in changes in local blood flow, which in turn result in changes to the local magnetic resonance. Magnetic resonance can be measured by applying a magnetic field to the brain and measuring consequent photon emissions (the principle upon which all magnetic resonance imaging (MRI) is based). Like EEG, fNIRS, and MEG, fMRI is non-invasive. Like MEG, however, it demands an extremely large magnet and a large, expensive sensor unit, with no obvious path to miniaturization. Furthermore, because fMRI depends on changes in blood flow, which lag behind electrical activity, the fMRI signal is both delayed and smoothed relative to the underlying brain activity, resulting in poor temporal precision (on the order of seconds). Consequently, fMRI is unlikely to play

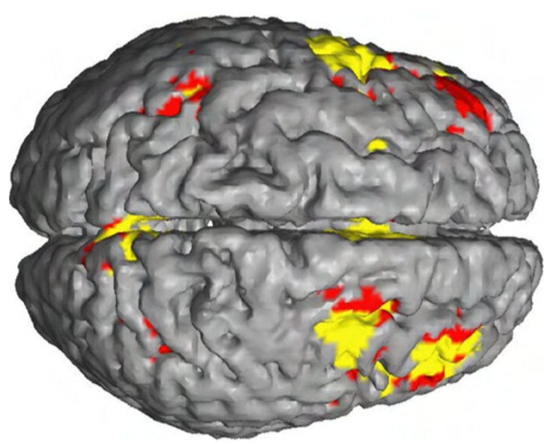

Fig. 2.19 fMRI can non-invasively monitor brain activity, but requires large, expensive equipment and offers poor temporal precision. Courtesy Tor Wager.

a role in always-available interfaces in the foreseeable future. However, fMRI continues to be a tremendously valuable research technology for studying the neural correlates of high-level cognition, yielding signals corresponding to prosody [112], language processing [63, 87], and object perception [17, 65]. fMRI therefore may impact the future of (potentially portable) BCIs, even if fMRI itself is constrained to research environments.

## 2.6 Muscle-Computer Interfaces

The previous section highlights the major challenge with BCIs as a supporting technology for always-available interfaces: though the electrical activity of the brain represents an appealing target for sampling a user's intent, the relevant signals are simply *too complex* and *too difficult to access* for practical direct-control applications right now. However, directly recording the electrical activity of a user's *muscles* represents an interesting intermediate: still a clear representation of a user's intent, and still measurable without requiring the user to hold a physical device in her hand, but much more accessible than the signals underlying BCIs.

When we initiate a voluntary motor action — for example, moving a limb or tensing our muscles without moving — the brain sends an electrochemical signal through the spinal cord. This signal is very similar to the signals brain cells use to communicate with each other, which is the signal sensed directly or indirectly by all the sensors discussed in the previous section. When this signal reaches the muscle, it continues to travel up and down the length of a muscle using a similar mechanism, and muscle cells respond by contracting. This signal can be measured as it propagates through the musculature by inserting electrodes through the skin and into the muscle, a measurement technique known as *electromyography* (EMG). However, for purposes of this review we assume that needle insertion is unlikely to be practical for consumer interfaces in the foreseeable future, due to concerns around both safety and comfort. Fortunately, the same signal can also be measured by placing electrodes on the surface of the skin, a measurement technique known as *surface electromyography* (sEMG) (Figure 2.20). sEMG senses electrical

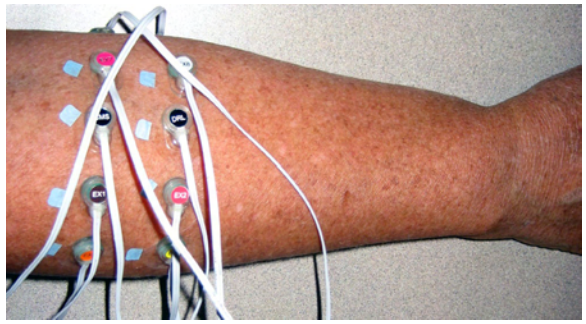

Fig. 2.20 Surface electromyography (sEMG) allows non-invasive measurement of electrical muscle activity [117]. © ACM 2008.

muscle potentials through metal electrodes placed against the skin, generally permanently plated in an electrolyte (often silver chloride) that enhances conductivity, along with a conductive gel that is applied each time the sensors are put on. sEMG provides lower signal amplitude and a lower signal-to-noise ratio than needle-based EMG, but provides a signal that is much higher-amplitude than that provided by, for example, EEG (millivolts instead of microvolts). This affords significant tolerance to environmental noise compared to electrical brain sensing. More importantly, the signals one observes through EMG or sEMG are complex, but *much* simpler than those observed through EEG. Roughly speaking, higher-amplitude EMG signals correspond to more muscle contraction, whereas an EEG signal collected from almost anywhere in the brain is a very complex function of perception, motor intent, high-level cognition, etc.

Another important factor supporting the plausibility of sEMG as an interface technology is the location of the musculature controlling the hands and fingers. Humans possess great dexterity in our hands, which is why most of the computer input devices we use today are designed for communication through our fingers. The muscles that control our hands and fingers (with the exception of the thumb) are located on the forearm, several inches away from the hand, connected to the skeleton of the hand by a complex system of tendons. This suggests that a computer input system could sense these muscles — and hence sense finger movements or intended finger movements — with an armband that does not constrain a user's ability to interact normally with the physical world through his fingers. The remainder of this section will look at several research projects that attempt to realize this vision. We do not survey the long history of sEMG in clinical applications, such as monitoring muscle progress during rehabilitation, but rather refer the reader to Ref. [126] for a survey of this space; similarly, we do not survey the application of sEMG for control of prosthetic devices, but rather refer the reader to Ref. [96]. Furthermore, a detailed discussion of the signal processing and machine learning techniques used in the work discussed is beyond the scope of this review; for reviews of these techniques, see Refs. [85, 108]. Instead, we focus on work applicable to consumer scenarios, and discuss the strengths and weaknesses of several research projects.

Saponas et al. [117] use a surface EMG sensors placed on the upper forearm to classify a user's finger movements via supervised machine learning. This work demonstrated that it was possible to discriminate among fingers tapped and on a surface and lifted off that surface using the sEMG signal, with classification accuracies in the vicinity of 75–90% for five- and six-class problems. However, this work suffered several practical limitations. The system assumed a relatively static hand (resting on a table). This work also relied on an expensive, large, wired apparatus, including impractical conductive gel (suitable for medical applications, but likely unsuitable for consumer applications). This apparatus required an experimenter to apply several sensing electrodes manually, a time-consuming process. Perhaps most significantly, this work assumed that a user would train a supervised learning system for several minutes prior to a classification session.

The same research group explored the application of similar techniques to a wider variety of scenarios in [115], which relaxed the restriction that a hand be held against a surface, and demonstrated the feasibility of in-air finger gestures where the hand is free to rotate or where the hand may be holding another object. This work still relied on a wired clinical EMG apparatus, but Saponas et al. [116] (Figure 2.21) relaxed this requirement by introducing a wireless device using dry electrodes that did not require careful placement of each sensing electrode. In this work, the authors also demonstrated that a user could use a

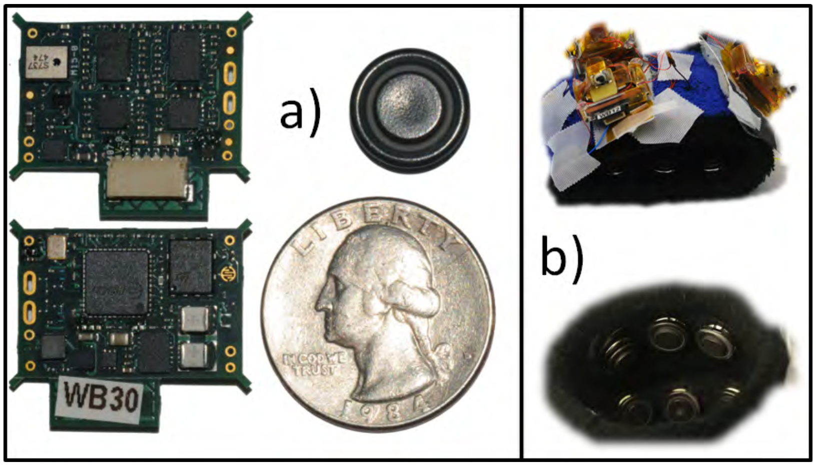


Fig. 2.21 Saponas et al. present a wireless, dry sEMG device [116]. © ACM 2010.

trained classifier even multiple days after training, where the sensor had been removed in the intervening time. However, even at this stage, this thread of research still relied on supervised classification: a user had to spend several minutes training the system prior to use, potentially prohibitive for consumer scenarios. Furthermore, classification accuracies are still far from perfect (100%), suggesting the need for further refinement of the signal processing and machine learning techniques underlying this work.

Kim et al. [67] achieve high classification accuracies by requiring the user to perform more coarse gestures (whole-hand movement instead of finger movement); they use the control of a remote-control car as a test application. This approach complements the work discussed above, but the need for large hand motions may prove problematic for subtle gesture execution in some scenarios, and offers a low ceiling on the system's gesture vocabulary. On the other hand, this work not only provides high accuracies, but also uses a very simple configuration of just three electrodes, highlighting an interesting space on the cost-vs-functionality curve for EMG input.

Costanza et al. [25] address several questions around the practicality of EMG input, by exploring the hardware design of a wireless EMG sensor in more detail than was presented in [116], which they

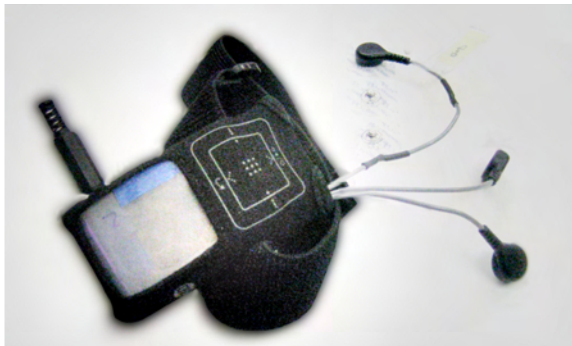

Fig. 2.22 Costanza et al. [25] integrate EMG electrodes into an armband form factor that might be plausible for consumer use. © ACM 2007.

incorporate into an armband that approaches a form factor that would be plausible for consumer use (Figure 2.22). Furthermore, this work explores the *visibility* (or rather, invisibility) of EMG-based gestures to an outside observer: by wearing an EMG sensor on an armband under long sleeves, a user is able to perform very subtle gestures, which the authors show are rarely visible even to an observer explicitly tasked with detecting gestures.

Additional ongoing work is attempting to further the state of the art in signal processing and classification for consumer EMG applications. Wheeler et al. [147] present a Bayesian method for extracting individual muscle activation signals from the ensemble activity sensed by sEMG electrodes, using knowledge of muscle physiology to offer a potentially richer feature set for classification than those used in the work discussed so far. Tenore et al. [138] present time-domain techniques and leverage a dense, 32-electrode array to achieve high accuracy in classifying finger movements. Ju et al. [62] address perhaps the two most challenging problems in this space through novel algorithms: the need for adaptation over time (to account for changes in the EMG signal when a device is worn for long periods of time) and the need for cross-user training which minimizes the burden on each user. After exploring a variety of static classification techniques akin to those used in the above work, this work explores adaptive stream processing for EMG signals and shows promising progress toward solving both of these problems.

## 2.7 Emerging Sensors

This survey has thus far focused on categories of sensors that have been relatively deeply explored, at least in the academic literature. In this section, we will turn our attention to several emerging sensor categories that have received less attention, but may represent promising approaches to always-available input.

### 2.7.1 Mechanical Sensing

The sensing of mechanical impulses traveling along or around the body represents one such sensing category. Harrison et al.'s *Skinput* system, for example, explores the use of piezoelectric accelerometers — worn in a compact armband that could be situated on the wrist or forearm — to classify the location and type of finger taps performed by one hand on the opposite arm [45] (Figure 2.23). They find that the mechanical impulses traveling up the arm vary enough among tapped locations

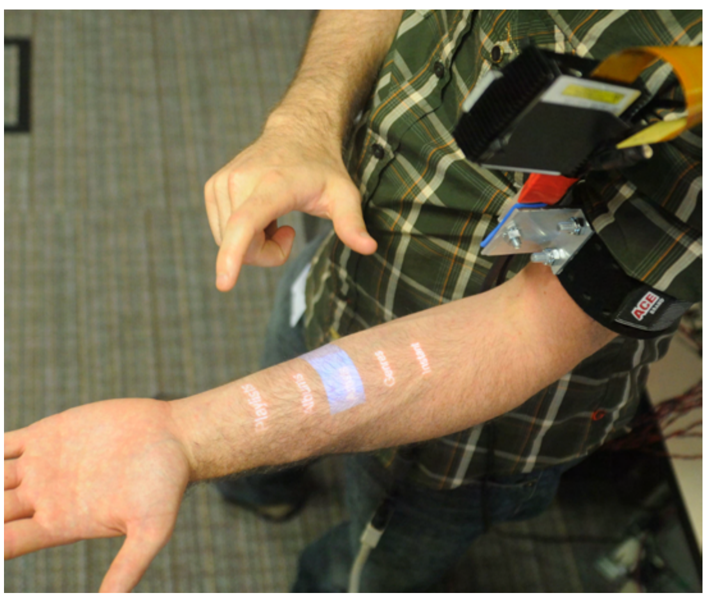


Fig. 2.23 Harrison et al.'s *Skinput* system combines piezoelectric accelerometers (worn in an armband) with a shoulder-mounted projector to prototype an on-body UI [45]. © ACM 2010.

that such classification is possible, up to as many as 10 unique locations on the forearm. This approach effectively turns the surface of the arm into a tap-sensitive surface. The authors combine their approach with a shoulder-mounted projector to highlight the possible use of this approach for bringing familiar interface paradigms such as buttons and scrolling menus to mobile scenarios. This approach also benefits from our ability to locate points on our bodies using our kinesthetic senses: if I ask you to close your eyes and tap one finger against the fingers of the opposite hand, for example, you will likely have no trouble doing so. This supports the feasibility of on-body, bi-manual interfaces like Skinput for mobile scenarios, although further validation of robustness is necessary.

A related approach to mechanical sensing is adopted by Amento et al. [3] (Figure 2.24), who use wrist-mounted microphones to classify gestures based on the unique sounds that propagate through the hand and arm when several finger gestures are performed: "tapping", "rubbing", and "flicking" gestures, for example, each generate a unique bioacoustic signature. Though this offers a smaller vocabulary than the approach taken by Harrison et al., it is a single-handed interaction technique that may be more subtle and may be practical for scenarios where bi-manual interaction is not.

### 2.7.2 Magnetic Sensing

The recent availability of magnetometers (sensors that report the orientation and strength of a local magnetic field), including their

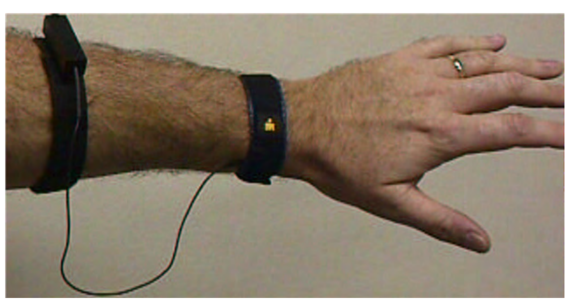

Fig. 2.24 Amento et al. use wrist-mounted microphones to classify gestures based on the unique sounds of various finger gestures [3]. Courtesy Brian Amento.


Fig. 2.25 Harrison and Hudson's *Abracadabra* system couples a wrist-mounted magnetometer with a passive magnet worn on the finger to provide three-dimensional input [44]. © ACM 2009.

incorporation into some mobile phones, has spawned some exploration of magnetic sensing for input. One important application is to add absolute orientation sensing to a system of accelerometers and gyroscopes, as used in [148]. However, other work explores the direct of magnetic sensing for input. Harrison et al.'s *Abracadabra* system [44] (Figure 2.25) couples a wrist-mounted magnetometer with a passive magnet worn on a finger of the opposite hand to provide three-dimensional input supporting a variety of interaction techniques. While this does require two physical components, one is easily conceived as a component in a wristwatch, the other in a ring, suggesting that this approach could be suitable for a variety of mobile interaction scenarios. Ketabdar et al. [66] explore a similar approach using the magnetometer built into a commercial mobile phone. In this case, the input signal is used specifically to control a computer-based musical instrument, highlighting the potential that many of the systems discussed here offer for on-the-go *creativity*. Askbrook et al. present *Nenya*, a finger-ring input device that uses magnetic tracking performed by a wrist-worn sensor [4]. In this system, users twist the magnetic ring on the finger for selection, and slide it along the finger for clicking. The authors propose that this provides fast access to analog input in a form factor that is socially acceptable, and their user studies explore both one- and two-handed interaction with the device.

### 2.7.3 Electrical Sensing

We previously discussed the use of electroencephalography and electromyography (using electrodes placed on the skin to monitor brain or muscle activity, respectively) — for computer input. Perhaps due to its long history in medical sensing and in prosthetics, these techniques have received quite a bit of attention. But several other types of electrical sensing have also begun to emerge as candidate approaches to always-available input. For example, Rekimoto [111] (Figure 2.26) demonstrates the incorporation of a unique capacitive sensor into a watch-like form factor, leveraging the observation that the wrist changes shape (and cross-sectional area) with different hand postures. This change in shape results in a change in capacitance, which — combined with

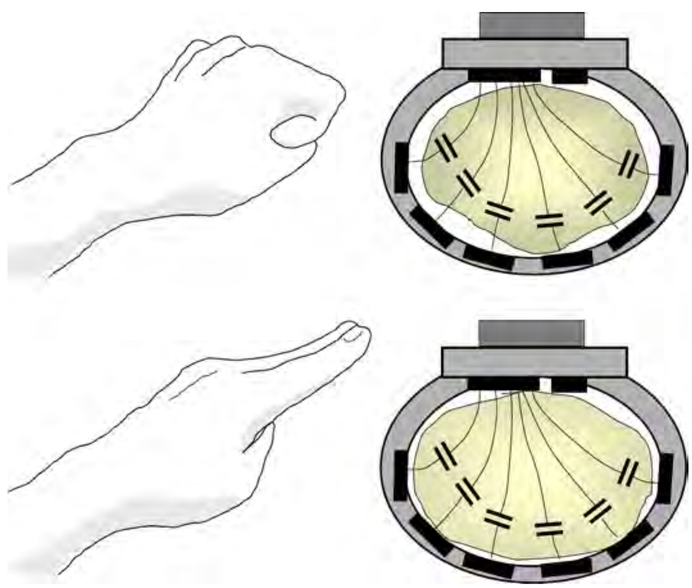

Fig. 2.26 Rekimoto's *GestureWrist* system [111] leverages the fact that the wrist changes shape — and therefore changes its electrical properties — when hand posture changes. © IEEE 2001.

additional information provided by an accelerometer — allows classification of a variety of hand gestures.

In a radically different approach to electrical sensing, Cohn et al. [24] observe that the human body not only *generates* electrical signals (as leveraged by sEMG), but also *captures* electrical signals radiating through the environment. In other words, the human body serves as a powerful antenna. Furthermore, they observe that a typical home contains a significant amount of electrical noise — in particular that this noise varies among locations within the home, due to the unique electrical signatures of appliances and wiring patterns. This work thus uses a body-coupled analog-to-digital converter to collect electrical noise in a home environment, and identifies variations in that "noise" to classify the locations through which a user is traveling, and even gestures that a user is executing. This approach offers an interesting hybrid between traditional gesture recognition that depends on instrumentation (e.g., cameras) and the always-available, on-body techniques discussed throughout this survey. Here the system depends only on the presence of location-specific noise in the environment, an assumption that is reasonable for a wide variety of scenarios. The generalization of their classification to novel environments, however, is left to future work, so this approach is still environment-dependent to some degree.

# 3

# Always-Available Output Technologies

So far, we have discussed a variety of *mobile input technologies*: sensors that capture some component of user intent for interpretation by a mobile computer. However, nearly every exchange a user has with a computer requires both input and output components. Output may be as simple as confirmatory feedback (e.g., a "click" to let you know that your photo was taken), but more often represents a more complex relaying of content or state from the machine to the user, often in a real-time, closed loop. The mechanism we most often rely on to deliver this information in non-mobile environments — and even in traditional mobile environments — is the pixel-based display. However, we argue that for a computing environment to be truly always-available, other feedback mechanisms will be necessary. Directing a user's visual attention to even a handheld screen violates two of our requirements for always-available computing: the act of redirecting your eyes away from the world is a significant cognitive disruption (requirement "1"), and a handheld screen is not typically available in hands-busy scenarios (requirement "4").

Current non-visual feedback channels on mobile devices — primarily audio feedback and simple vibration — do not provide nearly adequate bandwidth to enable interaction in scenarios where the visual channel is unavailable, nor are they designed to do so. Consequently, although this review's main focus is on input technologies, we devote this section to major trends in mobile *output* that will be important to realizing the vision of always-available computing.

## 3.1 Haptic Feedback

Whether we are typing on a keyboard, pressing a button, or even writing with a pencil, our brains receive a constant stream of touch sensations that play an important role in manipulating tools. As such, consumer electronics make use of our sense of touch not only through the mechanical design of objects but also through programmable haptic output. Piezoelectric vibration elements and off-center weighted motors have achieved significant commercial success in mobile phones and game controllers, respectively, but a variety of other actuators have been employed to create haptic sensations in research environments. Iwamoto et al., for example, employ ultrasound waves for contact-free haptic stimulation [59], and Bau et al. employ electrovibrations to create textures on a touch screen [10].

While some of the most commercially successful uses of haptics require a handheld instrument (e.g., a game controller or a mobile phone), haptic output is appealing as an always-available feedback mechanism because it *can* be applied away from our hands in a portable form (e.g., around a watch band). In this section, therefore, we explore the applicability of haptic feedback to always-available interaction.

Several groups have developed prototype belts that use vibratory elements to indicate direction and assist a user with navigation [141, 142]. One could imagine employing this approach to guide a user to the closest coffee shop, without interrupting the user's conversation with a friend. These examples fit into a more general category of using haptics for ambient output around the body [79]. In addition to belts, researchers in this area have explored a wide variety of form factors including vibrotactile actuators built into the shoulder pads of clothes [139] and arrays of vibrotactile actuators that can "draw" patterns on a person's arm [21]. Researchers have also attempted to characterize what types of vibrotactile output can be successfully interpreted by a person. For example, Chen et al. found that people cannot easily distinguish nearby vibrotactile actuators; when placing a $3 \times 3$ grid on both the top and bottom of the wrist, they found that people's ability to distinguish which one of the 18 tactors vibrated ranged from 30% to 73% depending on the location [19]. However, they do point out that participants could identify which side of their wrist the vibration came from 93% of the time. In a related study, Oakley et al. observed similar results for localization and also noted that people are better at distinguishing linear change in location around their arm (like a watch band) than along the length of their arm [94] (Figure 3.1). One of the main practical barriers to commercializing these approaches is finding wearable form factors that comfortably accommodate computation, communication, and actuators without negatively impacting comfort, durability, and washability of garments or worn accessories.

In addition to the above examples of haptic output being worn on the body, researchers have also explored techniques for haptic output

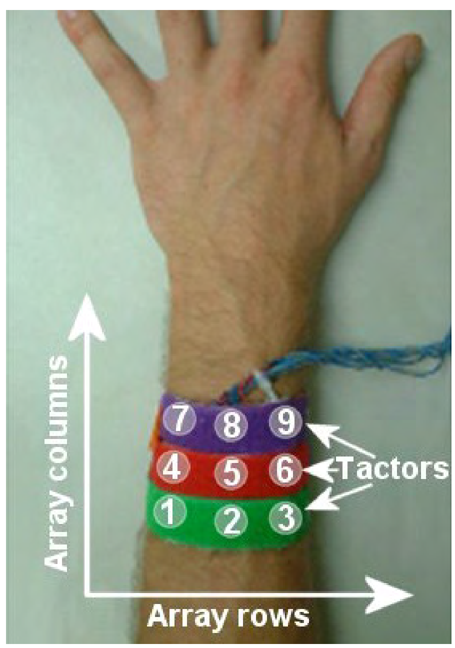


Fig. 3.1 Oakley et al. [94] use a $3 \times 3$ array of haptic actuators to explore perceptual questions around a plausible form factor for wearable haptics. © IEEE 2006.

on mobile devices [47]. These techniques focus on creating spatial and temporal patterns that we can perceive with our *fingers* when grasping a device. Touch receptors are packed more densely on our fingers than elsewhere on our bodies; as a result, haptic interfaces targeting the hands and fingers can potentially convey more bits of information in a shorter period of time, using less space. For example, Chang et al. [95] explore the use of haptics for communication, presenting a device with 12 vibrotactile actuators distributed across four fingers. This approach might enable a user to "feel" who is calling his/her phone, without needing to pull the phone out and glance at the display. This form factor holds promise for always-available interfaces, but requires a device to be held in the user's hand, and thus may not be appropriate for all always-available scenarios.

## 3.2 Audio Feedback

Nearly all mobile devices provide audio output, used ubiquitously for phone calls, listening to music, and playing games. This ubiquitous availability, combined with the increasing prevalence of always-available headsets (e.g., Bluetooh earpieces), suggests that audio may indeed play a role in always-available interaction. Furthermore, smaller wireless earpieces are emerging that can fit invisibly inside a person's ear canal, easing the social awkwardness of visibly wearing headphones while interacting with other people.

While audio feedback can be *available* at all times, it is only effective as an always-available output mode if it does not impede a person's primary tasks. However, the use of language in computer interfaces creates the potential for cognitive interference when the user may also be engaged in language-centric real-world tasks, a problem discussed above with respect to speech input. For example, it is difficult for a person to listen to spoken language in an earpiece while also engaged in conversation. Although it is possible for a person to become skilled at simultaneous listening and conversing, it is still a significant challenge to incorporate linguistic audio into mobile interfaces that a person would use while also using engaged in linguistic tasks such as reading or conversation.

Consequently, researchers have also explored several opportunities for non-linguistic mobile audio feedback. One of these themes is passive support of spatial navigation tasks. Holland et al. demonstrate the *AudioGPS* technique for generating tones to indicate bearing and distance for pedestrian navigation [48]. This approach conveys the bearing of a destination relative to the user by manipulating the perceived location of a synthesized tone, and indicates distance to that destination by varying the rate at which those tones are generated. This leverages our ability to robustly perceive the location of sounds presented in headphones; Vazquez-Alvarez and Brewster have demonstrated that people can spatially discriminate among five audio sources over a 180-degree range using typical headphones [143]. Sodnik et al. have even shown that a spatial audio interface can be less distracting than a visual interface for driver navigation in vehicles [129]. In addition to these navigation tasks, Li et al.'s *BlindSight* system [75] demonstrates that non-linguistic audio can also be used for quick, eyes-free querying of a person's calendar even while they are engaged in a phone call. As these research projects demonstrate, the availability of small, wireless headphones — combined with new non-linguistic audio techniques — suggests that audio feedback may offer significant value for always-available mobile interfaces.

## 3.3 Glasses and Other Mobile Displays

The primary output mechanism in almost all computing tasks has traditionally been visual, due to the high bandwidth that visual displays provide relative to audio or haptic displays. This bandwidth is critical for always-available interfaces that aim to minimize the duration of interruptions. For example, if we are alerted to the arrival of an email, a visual display enables us to quickly skim the contents of that email. Visual displays also have the ability to convey non-linguistic information quickly, through images and video. In addition to these high-bandwidth applications of visual output, a low-bandwidth but useful property of our visual system is our ability to perceive shapes, color, and motion through our peripheral vision while keeping our primary visual attention on another task. In an always-available interface, this could enable a display to keep us apprised of information (e.g., incoming messages, weather, direction and distance to nearest coffee shop) without interfering with our primary task. From these properties, we can broadly say that visual displays are a rich output medium with the potential to provide fast transitions in and out of the interface (as fast as a glance), while minimally interfering with our primary tasks. However, the practical challenge for always-available visual output is building displays in a form factor that is portable, comfortable, and socially acceptable. Below, we review the state-of-the-art in mobile visual output.

### 3.3.1 Glasses

Wearable computing has long sought an effective mobile display built into eyeglasses. This is a somewhat natural choice given that eyeglasses are a commonly-worn accessory and perhaps have enough bulk to hide a display's components. A simple version of this vision is an LCD-based display that clips onto one side of a pair of eyeglasses[1] (Figure 3.2).

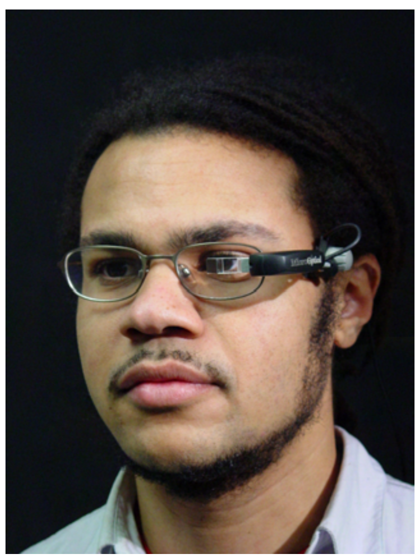

Fig. 3.2 The Teleglass project provides an LCD display that clips on to any pair of eyeglasses. Courtesy Hrvoje Benko and Alex Olwal.

[1] Arisawa Teleglass, publicly reported pre-production unit, http://www.arisawa.co.jp/en/.

Fig. 3.3 Progress in optical technologies allow displays, such as this prototype from Lumus Ltd., that present information on transparent glass that does not obstruct the wearer's view. Courtesy Lumus Ltd.

The main drawback of these displays is that they are bulky and partially obstruct the wearer's vision, even when powered off. More recent prototype eyeglasses go farther toward realizing clear, display-equipped eyeglasses[2,3] (Figure 3.3). These prototypes are heavier than traditional eyeglasses, but only minimally obstruct a user's vision. If this technology can continue to improve, we believe that it is likely the best candidate for always-available output in the near-term future.

### 3.3.2 Contact Lenses

Saeedi et al. have taken the concept of a ubiquitous display in front of the eyes to an even more invisible level by creating initial prototypes of LED-array-based displays built into contact lenses [129] (Figure 3.4). This technology is in the early stages of development, requiring many more advances before it is ready for human use. However, even if future contact lenses are only able to display a line of text and a few colored dots in the periphery, they would, in many ways, be the ultimate always-available output technique.

### 3.3.3 On-Body Projection

Another approach to creating quickly accessible mobile displays is to put a display directly on a person's body. For example, researchers have

[2] Microsovision Wearable Displays, Microvision, Inc., http://www.microvision.com/.
[3] Lumus Personal Displays, Lumus [114] Ltd., http://www.lumusvision.com/.

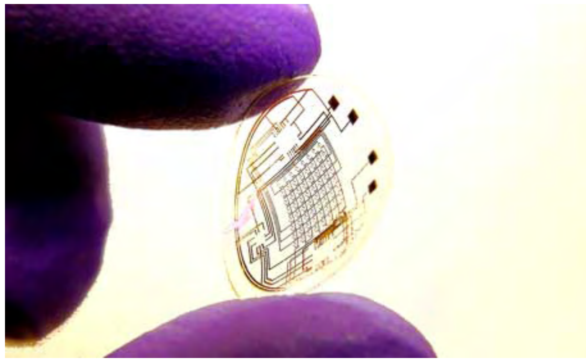

Fig. 3.4 Saeedi et al. are working toward embedding displays directly on contact lenses. Courtesy Babak Parviz.

explored projecting displays directly onto a person's arm or hand, coupled with input techniques for detecting direct interaction with the projection [45, 86] (Figure 2.23). On-body projection is appealing because it does not require users to wear special purpose glasses or contact lenses, and — unlike glasses or contact lenses — on-body projection also offers the potential of a *shared* portable display. The main drawback of wearable projection is the challenge of creating a projector that is easily worn on the body and provides a bright enough projection to be seen in common lighting conditions (e.g., daylight). Even the smallest current hand-held projectors (often referred to as "pico projectors") are likely too large for wearable applications and are not bright enough for practical use in mobile environments: today's best devices are on the order of five cubic inches and offer only 30 lumens of brightness.[4] For comparison, 1000 lumens is generally considered the bottom end of suitability for projection in an office environment, where lights might be on and the projection surface might be several feet away and several feet tall, and 2000–5000 lumens is typical for desktop or ceiling-mounted projectors.

[4] 3M MP180, 3M, http://solutions.3m.com/wps/portal/3M/en_US/Pocket/Projector/Main/Products/MP180/.

### 3.3.4 Clothing-Based Displays

On-body displays need not be projected; they can also be embedded in our clothes. The main drawback of this approach is the requirement that the clothes we wear every day have display technology embedded within them. This may be incompatible with traditional properties of clothing such as washability, durability, disposability (i.e., low cost), and flexible aesthetics. There are, however, several plausible technologies for in-clothing displays. On obvious approach is to embed LEDs directly into fabrics [15], allowing a familiar paradigm of turning individual photo-elements on or off to create images. More generally, a variety of *electroluminescent* elements (anything that emits light when electrical current passes through it) are available today. However, the manufacturing of embedded LEDs or electroluminescent patterns that can be washed and mechanically protected, however, is challenging. Heat-sensitive dyes that change colors with variations in temperature, or *thermochromics*, provide a potential alternative: embed dye in fabric using traditional processes, and use wires or other elements to control the temperature of these dyes, possibly without mechanical contact. Although this technique has been used for decades to produce clothes or accessories that change color when heated as a result of being touched by a human hand, this approach is at its early stages in terms of manufacturing clothing whose appearance can be computer-controlled.

# 4

# Challenges and Opportunities

In this review, we have surveyed a relatively broad swath of emerging input technologies that we believe will be instrumental in enabling always-available mobile interaction. In this section, we discuss several higher-level challenges that span many of these technologies. As advances in sensors and materials continue to drive all of the technologies we have discussed so far, it is the areas discussed in this section that we believe offer human-computer interaction researchers opportunities for the broadest impact on always-available interaction.

## 4.1 Systematically Handling Ambiguity

Most traditional input devices have been designed to provide a stream of data that is as well-defined as possible. For example, there is little ambiguity on whether or not a key on the keyboard has been pressed, or how much the mouse has moved on a surface. However, many of the newer modalities described in this review tend to infer action and intent from sensors that produce much noisier raw signals. Although just about every researcher working on new input modalities attempts to remove ambiguity as best they can, recognition errors for some of these modalities will likely remain an intrinsic part of the sensing process and will never be completely eliminated. Hence, we believe that we must systematically handle — or better yet, design for — the two main classes of ambiguity: *recognition ambiguity* and *segmentation ambiguity*.

Researchers have used multiple techniques to reduce recognition errors. Because of its maturity as a field, a good bit of this work has been done in the speech recognition domain, but results are often relevant and applicable to newer modalities as well. For example, early work on speech recognition explored ways of providing appropriate feedback for error correction [2]. In different domains, researchers have found that appropriate feedback allows users to form mental models of the system, and actually helps them perform gestures that can be better recognized (e.g., [115]). While we will not document this literature in detail, there have also been many efforts to utilize multimodal interfaces in order to reduce ambiguity and improve recognition accuracy. For a survey of the literature in this field see Refs. [30] and [98].

Researchers have also worked on handling recognition ambiguity by providing correction mechanisms that allow users to quickly and cheaply roll back and re-specify the intended action. Shilman et al. utilize past handwriting input associated with an error as well as the user's correction of that error, in combination with a set of gestures that allow the user to further assist the recognizer, to improve on recognition correction [127]. Similarly, Mankoff et al. perform a survey of error correction techniques and find that they fall into two basic categories: repetition and choice [80]. They develop the *OOPs* toolkit, and a set of associated interaction techniques, to support resolution of input ambiguity.

The second class of ambiguity in always-available interfaces is *segmentation ambiguity*. Since, we are claiming that the input modality is "always-available" and since the user does not always intend to be interacting with the computer, especially as they go about their real-world tasks, the system must be smart enough to distinguish between actions in the real world and explicit commands to the system. The confusion between the two is often referred to in the eye-tracking and gesture tracking literature as the "Midas Touch problem". Huckauf et al. develop an eye-tracking-based input system based on explicit gestures that are unlikely to naturally occur for actuating commands [51]. Rather than utilizing dwell time, as is usually done in eye-tracking control applications, they suggest anti-saccades, or quick glances, at a copy of an object to specify intent to operate on that object. In their *Snap Clutch* work, Istance et al. provide a lightweight mechanism to turn on/off parts of the control mechanism (namely, gaze-based cursor control) while maintaining some amount of continued input (selection) [58]. In their work on muscle-computer interfaces, Saponas et al. propose explicit actuation gestures as well as a combination of gestures that do not usually occur naturally in everyday tasks in order both to circumvent the Midas Touch problem and to increase effective recognition accuracy [115]. In general, we believe there are opportunities for better segmentation of naturally occurring gestures and explicit ones, as well as more systematic approaches to defining gesture languages around on/off mechanisms.

More generally, there has been effort to model uncertainty and handle it as a normal and expected part of the input process. Starting in the early 1990s, Hudson and Newell proposed the notion of probabilistic state machines that model uncertainty and maintain assessments of the probabilities for alternate means of gestures [52]. They claim that doing so allows the system to make more informed decisions about when to invoke actions, thus leading to more robust performance. In follow up work, Schwartz et al. develop a toolkit (and some very clever thinking) around how ambiguity in input could be passed into higher levels of the UI [123]. For example, a Web form designed for uncertain input focus (perhaps expecting text or speech entry) could evaluate each possible text box the user might be typing into, and place the input in the box whose input model best fits the content. They present multiple prototype interfaces and applications for this model and argue that a fundamentally new computing paradigm will have to be designed as we continue to evolve our interaction techniques to ones that include more and more ambiguity.

Despite the extensive research done in algorithmically minimizing ambiguity and developing interface metaphors around ambiguity, this remains a tremendous source of interface breakdowns even today: input has become more ambiguous much faster than interface metaphors have evolved to accommodate this trend. In fact, interacting via of our increasingly-ambiguous input devices still relies on metaphors developed for their unambiguous ancestors (the keyboard and mouse). For example, touch-based input devices still depend on users clicking individual points (e.g., to click on buttons or links), a task that is quite straightforward with a mouse but quite ambiguous with a finger that spans literally hundreds of pixels and possibly dozens of potential input targets. Consequently, misclassified touches frequently create user frustration, or — more subtly — limit the density of information that devices can present to users. This suggests an important opportunity for the HCI community: the development of metaphors that leverage the multi-dimensional nature of touch as a means to compensate for the ambiguity of finding a single "touch point".

As another example, speech-based search is available on most mobile devices, but still relies largely on a familiar paradigm of transcribing speech into text (an ambiguous process) and executing a discrete (and potentially incorrect) query, perhaps offering the user a chance to correct that query. The use of implicit or explicit *context* to resolve an inherently-ambiguous speech query represents an exciting area of exploration, one which will require collaboration between HCI and speech researchers.

## 4.2 Sensor Fusion

Having described many emerging interaction technologies, one may reasonably ask the question of which will is likely to be the ideal modality, the "mouse and keyboard" of next-generation computing devices. While we believe that this is a reasonable question, we do not believe that there will exist a single solution. The shift from a well-defined and rather static computing environment in the desktop computing world to the dynamic and ever-changing scenarios in the mobile computing world will likely necessitate a combination of modalities working in close complement. More importantly, given the ambiguity we are introducing in many of our new modalities and the fact that multimodal and multi-sensor modalities have been shown to improve robustness and add richness to the interactions, we believe creative sensor fusion will be a large topic of interest as we move forward. As described in previous sections, many projects are already starting to explore this (e.g., [102, 111, 144, 151]), but we believe work explicitly aimed at more systematic sensor fusion will be important and grow significantly in this domain.

## 4.3 Gesture Design and Usability

There tend to be several phases in the development of any new interaction modality. The early phase is typically proof-of-concept: a developer or designer sets out to determine whether a new sensing technology works at all. Applications and particular use cases tend to be relatively ill-defined, and the pure novelty of the technology itself drives research goals and reader attention. As the technology matures, it is often used to emulate existing modalities. For example, touch- or gesture-based interaction modalities are often used to emulate mouse and keyboard interaction and applied to windows, icon, menu, pointer (WIMP) interfaces. Finally, at full maturity, we see specific affordances, applications, and paradigms tailored to take advantage of the properties of the modality. We believe that many of the modalities treated in this review are in transition from emulation to maturity, and that researchers continue to push hard on design of appropriate interaction techniques as well as usage and learning affordances, all of which are also well served (i.e., recognized) by the particular technology.

For instance, researchers working on gesture recognition concern themselves with systems and tools that allow developers to design gesture sets that are both easy for users to execute and learn, but that also make it tractable for the sensors and computers to differentiate and recognize. Long et al. present tests they ran using their pen-based gesture design tool describing how developers do not tend to understand the nuts and bolts of recognition engines and must be guided as they design gesture sets [22]. They also found the need for support in iteratively testing these gesture sets. Building on this, Ashbrook and Starner built *MAGIC*, which extends these findings to support motion gesture design [7]. Implicitly (and sometimes explicitly), many of these tools consider many important factors, including social acceptance and cultural appropriateness, learnability and memorability, the ability of the system to differentiate and recognize gestures, and to a smaller degree so far, the fit of the modality to its applications, and vice versa.

OctoPocus combines "feed-forward" mechanisms (pre-gesture help, guides, animations) as well as feedback (post gesture recognition results) to help users learn and remember gesture sets, and techniques surrounding this [9]. The exercise suggests how difficult it is to design new modalities, systems, and applications, and points at the need for a range of design methodologies and principles in doing so. Clearly, there is much work to be done in this space, and we hope that this review provides some of the basis for new researchers in the area to identify new problems and approaches and to innovate in the way we design new modalities, gestures, and applications for these.

Many researchers have specifically explored the universality of gestures and whether or not factors like culture affect execution and memorability. This is especially important since the mobile computing task is often embedded in the real-world around other people, sometimes even involving the other people. Rehm et al. use the Nintendo Wii controller to input accelerometer gestures and find multiple cultural differences, even down to their resting poses [110]. This suggests the need to be sensitive to tuning recognition technologies, but also awareness and sensitivity so that we design interactions that are not socially awkward. That said, we also believe that no gesture set will ever be intrinsically "natural" and that users will always have to learn some part of the interaction. This has been true of all our modalities, even ones that we eventually consider to be second nature, like the mouse and keyboard. This is captured nicely by Stern et al., who present an analytic approach to designing gesture vocabularies by decomposing the problem into system constraints and user constraints and optimizing gesture sets for the overall utility [136].

Other researchers looking at cultural effects have focused on social acceptance of the gestural interfaces. Calkin et al. examine definitions of social acceptance, not only for the user but also for observers [88]. They identify factors such as user type (i.e., where they sit in the adoption curve), culture and time, as well as the actual interaction type. One of the findings they propose is that a reasonable amount of social acceptance is derived from the user's perception of others' ability to interpret the manipulations. This is interesting as it suggests that widespread adoption, or at least understanding, of any technology will lead to natural social acceptance. This is not only because "everyone is doing it," but just as much because everyone actually understands and can mentally attribute gestural actions to computing ones. Sensitive to this in their work on "Intimate Interfaces," Costanza et al. not only introduce motionless gestures sense through electromyography, as we have describe earlier in this review, but also assess how noticeable they were to informed observers [25]. They found that even people looking out had trouble identifying when a gesture was performed, which they concluded was a positive property of this modality.

Apart from the gestures themselves, there are pragmatic issues in many of these modalities, which tend to include sensors and devices that are worn on the body. While it is not within the scope of this review to discuss all the work in the wearable computing space, readers may see any of a number of surveys and published work in this space (e.g., [131, 132] as well as work from academic venues such as the International Symposium on Wearable Computers). Apart from the general work going on in the wearable computing domain, Ashbrook et al. investigate placement and user mobility on the time required to access an interface worn on the body [6]. They found that placing the device in a holster or pocket drastically increases (up to 78%) access time as opposed to a device that was, for example, always-available on the wrist. They suggest careful consideration of seamless access so users can most effectively compute on the go.

## 4.4 Cognitive Interference and Computational Senses

While much of the academic work today is conducted in relatively controlled laboratory (and sometimes field) settings, the vision we paint is one of infusing seamless computing into our everyday lives. There is much work to be done on exploring and improving the effectiveness of these modalities as they are used in the real world.

As articulated by Shneiderman in his review of the limits of speech recognition [128], for example, speech input often interferes with other speech-based interactions (e.g., human–human ones), and worse yet interferes significantly with other cognitive tasks. He describes why after 30 years of trying to provide airplane pilots with speech interfaces, complex functionality remains built into mechanical controls, as the cognitive load associated with speech and the conflicts it creates with the complex task at hand are too expensive. This is true of many of the examples discussed in this review. As another good example, most current brain–computer interface systems are designed for the user to invoke explicit thoughts in order to control some interaction. While controlling a computer with thought alone is impressive and inspiring, further work is required to evaluate the cognitive resources required for the task itself, and how the interaction can be designed so as to reduce cognitive interference.

Cognitive interference is of course neither new nor unique to interaction techniques. In fact, this is an area that has been studied at great length in the cognitive psychology and cognitive science fields (see [118] for an overview), and has become so important that the entire "cognitive ergonomics" subfield has grown up around it.

This field has a great deal to offer HCI: we believe that systematically understanding the physical and cognitive costs and benefits of various interaction methodologies is critical in designing interaction methods that allow us to integrate computer use while performing everyday tasks. We also believe that the nature of interference will necessitate the creation of an interaction ecosystem that is sensitive to the demands of various scenarios.

One ambitious goal for always-available interfaces is for them to be so unobtrusive and well integrated into our mental processes that they virtually function as another sense. Put another way, when do you "just know" whatever the computer is trying to tell you? Imagine having the ability to "feel" the presence of available WiFi , always "know" which direction is home (or the nearest coffee shop), or even "see" through walls and around buildings. Posed as a research question: can such senses be created through the unused bandwidth of our existing five senses combined with the power of mobile computing?

In this vein, several researchers have experimented augmenting human sensory perception. In 1947, Kohler fitted a subject with a special pair of glasses utilizing mirrors to present the eyes with an inverted image of the world [70]. He observed that despite the image inversion of the glasses, after several days the subject adapted to the visual distortion and began seeing right side up. Upon removing the glasses, the individual had sensations of the world being inverted. This early work reveals the flexibility of our sensory system.

More recently, researchers have explored adding or substituting senses. Bach-y-Rita et al. utilized the tongue as a human-input channel for sonar-like vision at night or for the blind [8]. And as discussed in Section 3.1, at least two groups have created belts that employ vibration to support navigation [141, 142]. These belts overload a person's "touch" sensation around the waist to actually create a new sense: a constant awareness of the recommended navigational strategy.

These examples illustrate that always-available interfaces have the potential to blur the line between an "application" and a "sense", yielding what we might refer to as a "super-human" experience. In the future, carefully-designed, always-available technologies might give us access to 100 kHz hearing, infrared to ultraviolet vision, and magnetic and electric field perception.

# 5

# Conclusion

In this review, we have presented the challenge that lies ahead of us in creating always-available computing interfaces. We assert that this forms the next large paradigm shift that will take us into the next generation of computing opportunities. We have laid out a starting point for properties of such interfaces and surveyed technologies that we believe may lead us closer to attaining the goal. While do not propose that these are a comprehensive set of building blocks required, we are impressed by the scope and depth of existing work, and hope that researchers continue not only to innovate in the space of sensors and techniques, but also to systematically solve some of the usability and design issues surrounding the integration of multiple input and output modalities in order to develop a richer mobile computing interface than we have ever known. The challenge is that simple, and also that difficult.

# References


[1] F. Ahmad and P. Musilek, "A keystroke and pointer control input interface for wearable computers," in *Proceedings of the International Conference on Pervasive Computing and Communications (PerCom)*, 2006.

[2] W. A. Ainsworth and S. R. Pratt, "Feedback strategies for error correction in speech recognition systems," *International Journal of Man-Machine Studies*, vol. 36, no. 6, pp. 833–842, June 1992.

[3] B. Amento, W. Hill, and L. Terveen, "The sound of one hand: A wrist-mounted bio-acoustic fingertip gesture interface," in *Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA)*, 2002.

[4] D. Ashbrook, P. Baudisch, and S. White, "Nenya: Subtle and eyes-free mobile input with a magnetically-tracked finger ring," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2011.

[5] D. Ashbrook, J. Clawson, K. Lyons, N. Patel, T. Starner, and S. Clara, "Quick-draw: The impact of mobility and on-body placement on device access time," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2008.

[6] D. Ashbrook, K. Lyons, and T. Starner, "An investigation into round touch-screen wristwatch interaction," in *Proceedings of the International Conference on Human Computer Interaction with Mobile Devices and Services (Mobile-HCI)*, 2008.

[7] D. Ashbrook and T. Starner, "MAGIC: A motion gesture design tool," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2010.

[8] P. Bach-y-Rita, C. C. Collins, F. Sauders, B. White, and L. Scadden, "Vision substitution by tactile image projection," *Nature*, vol. 221, pp. 963–964, 1969.

[9] O. Bau and W. E. Mackay, "OctoPocus: A dynamic guide for learning gesture-based command sets," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2008.

[10] O. Bau, I. Poupyrev, A. Israr, and C. Harrison, "TeslaTouch: Electrovibration for touch surfaces," in *Proceedings of the Annual ACM Symposium on User Interface Software and Technology (UIST)*, 2010.

[11] P. Baudisch and G. Chu, "Back-of-device interaction allows creating very small touch devices," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2009.

[12] T. J. Bradberry, R. J. Gentili, and J. L. Contreras-Vida, "Reconstructing three-dimensional hand movements from noninvasive electroencephalographic signals," *Journal of Neuroscience*, vol. 30, no. 9, pp. 3432–3437, 2010.

[13] L. Bretznerm, I. Laptev, and T. Lindeberg, "Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering," in *Proceedings of the International Conference on Automatic Face Gesture Recognition (FG)*, 2002.

[14] J. S. Brumberg, A. Nieto-Castanon, P. R. Kennedy, and F. H. Guenther, "Brain-computer interfaces for speech communication," *Speech Communication*, vol. 52, no. 4, pp. 367–379, Silent Speech Interfaces, April 2010.

[15] L. Buechley and M. Eisenberg, "Fabric PCBs, electronic sequins, and socket buttons: Techniques for E-textile craft," *Journal of Personal and Ubiquitous Computing*, vol. 13, no. 2, pp. 133–150, 2009.

[16] B. Buxton, "Multi-touch systems that I have known and loved," Originally written 2007, updated and retrieved 21 March 2011.

[17] K. M. Chang, T. M. Mitchell, and M. A. Just, "Quantitative modeling of the neural representations of objects: How semantic feature norms can account for fMRI activation," *Neuroimage*, [Epub ahead of print], May 2010.

[18] R. Chang, F. Wang, and P. You, "A survey on the development of multi-touch technology," in *Proceedings of the Asia-Pacific Conference on Wearable Computing Systems (APWCS)*, 2010.

[19] H.-Y. Chen, J. Santos, M. Graves, K. Kim, and H. Z. Tan, "Tactor localization at the wrist," in *Proceedings of the International Conference on Haptics: Perception, Devices and Scenarios (EuroHaptics)*, 2008.

[20] I. Cho, J. Sunwoo, Y. Son, M. Oh, and C. Lee, "Development of a single 3-axis accelerometer sensor based wearable gesture recognition band," in *Proceedings of the International Conference on Ubiquitous Intelligence and Computing (UIC)*, 2007.

[21] R. W. Cholewiak and A. A. Collins, "The generation of vibrotactile patterns on a linear array: Influences of body site, time, and presentation mode," *Perception and Psychophysics*, vol. 62, no. 6, pp. 1220–1235, 2000.

[22] A. Christian Long, Jr., J. A. Landay, and L. A. Rowe, "Implications for a gesture design tool," in *Proceedings of the International Conference on Human Factors in Computing Systems: The CHI is the Limit (CHI)*, 1999.

[23] J. Cohen, "Embedded speech recognition applications in mobile phones: Status, trends, and challenges," in *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008.

[24] G. Cohn, D. Morris, S. Patel, and D. Tan, "Your noise is my command: Sensing gestures using the body as an antenna," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2011.

[25] E. Costanza, S. A. Inverso, R. Allen, and P. Maes, "Intimate interfaces in action: Assessing the usability and subtlety of EMG-based motionless gestures," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2007.

[26] C. D. Cowper-Smith, E. Y. Lau, C. A. Helmick, G. A. Eskes, and D. A. Westwood, "Neural coding of movement direction in the healthy human brain," *PLoS One*, vol. 5, no. 10, pp. e13330, PMID: 20967197, October 2010.

[27] B. Denby, T. Schultz, K. Honda, T. Hueber, J. M. Gilbert, and J. S. Brumberg, "Silent speech interfaces," *Speech Communication*, vol. 52, no. 4, pp. 270–287, April 2010.

[28] H. Drewes, A. De Luca, and A. Schmidt, "Eye–gaze interaction for mobile phones," in *Proceedings of the International Conference on Mobile Technology, Applications, and Systems and the International Symposium on Computer Human Interaction in Mobile Technology (Mobility)*, 2007.

[29] A. T. Duchowski, "A breadth-first survey of eye tracking applications," *Behavior Research Methods, Instruments, and Computing*, vol. 34, no. 4, pp. 455–470, 2002.

[30] B. Dumas, D. Lalanne, and S. Oviatt, "Multimodal interfaces: A survey of principles, models and frameworks," in *Human Machine Interaction,* Lecture Notes in Computer Science, vol. 5440, (L. Kohlas, ed.), pp. 3–26, 2009.

[31] P. Eskandari and A. Erfanian, "Improving the performance of brain-computer interface through meditation practicing," in *Proceedings of the International Conference of the Engineering in Medicine and Biology Society (EMBC)*, 2008.

[32] M. Fukumoto and Y. Suenaga, ""FingeRing": A full-time wearable interface," in *Conference Companion on Human Factors in Computing Systems*, 1994.

[33] F. M. Galán, M. Nuttin, E. Lew, P. W. Ferrez, G. Vanacker, J. Philips, and J. D. R. Millán, "A brain-actuated wheelchair: Asynchronous and non-invasive brain–computer interfaces for continuous control of robots," *Clinical Neurophysiology*, vol. 119, no. 9, pp. 2159–2169, 2008.

[34] A. Girouard, E. Treacy Solovey, L. M. Hirshfield, K. Chauncey, A. Sassaroli, S. Fantini, and R. J. Jacob, "Distinguishing difficulty levels with non-invasive brain activity measurements," in *Proceedings of the International Conference on Human–Computer Interaction (Interact)*, 2009.

[35] M. Goto, K. Itou, and S. Hayamizu, "Speech completion: On-demand completion assistance using filled pauses for speech input interfaces," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 2002.

[36] M. Goto, K. Itou, and S. A. Hayamizu, "A real-time filled pause detection system for spontaneous speech recognition," in *Proceedings of the European Conference on Speech Communication and Technology (EUROSPEECH)*, 1999.

[37] M. Goto, K. Itou, K. Kitayama, and T. Kobayashi, "Speech-recognition interfaces for music information retrieval: "Speech completion" and "speech spotter"," in *Proceedings of the Conference of the International Society for Music Information Retrieval (ISMIR)*, 2004.

[38] M. Goto, K. Itou, and T. Kobayashi, "Speech interface exploiting intentionally-controlled nonverbal speech information," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST),* (Demos), 2005.

[39] D. Grimes, D. S. Tan, S. E. Hudson, P. Shenoy, and R. P. N. Rao, "Feasibility and pragmatics of classifying working memory load with an electroencephalograph," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2008.

[40] S. Gustafson, D. Bierwidth, and P. Baudisch, "Imaginary interfaces: Spatial interaction with empty hands and without visual feedback," in *Proceedings of Annual Symposium on User Interface Software and Technology (UIST)*, 2010.

[41] J. Y. Han, "Low-cost multi-touch sensing through frustrated total internal reflection," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2005.

[42] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478–500, March 2010.

[43] S. Harada, T. Scott Saponas, and J. A. Landay, "VoicePen: Augmenting pen input with simultaneous non-linguisitic vocalization," in *Proceedings of the International Conference on Multimodal Interfaces (ICMI)*, 2007.

[44] C. Harrison and S. E. Hudson, "Abracadabra: Wireless, high-precision, and unpowered finger input for very small mobile devices," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2009.

[45] C. Harrison, D. Tan, and D. Morris, "Skinput: Appropriating the body as an input surface," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2010.

[46] L. M. Hirshfield, E. T. Solovey, A. Girouard, J. Kebinger, R. J. K. Jacob, A. Sassaroli, and S. Fantini, "Brain measurement for usability testing and adaptive interfaces: An example of uncovering syntactic workload with functional near infrared spectroscopy," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2009.

[47] E. Hoggan, S. Anwar, and S. A. Brewster, "Mobile multi-actuator tactile displays," in *Proceedings of the International Conference on Haptic and Audio Interaction Design (HAID)*, 2007.

[48] S. Holland, D. R. Morse, and H. Gedenryd, "AudioGPS: Spatial audio navigation with a minimal attention interface," *Personal Ubiquitous Computing*, vol. 6, no. 4, pp. 253–259, January 2002.

[49] C. Holz and P. Baudisch, "The generalized perceived input point model and how to double touch accuracy by extracting fingerprints," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2010.

[50] J. Hong, J. Park, S. Han, and M. Hahn, "Sporadic noise reduction for robust speech recognition in mobile devices," in *Proceedings of the International Conference on Consumer Electronics (ICCE)*, 2011.

[51] A. Huckauf, T. Goettel, M. Heinbockel, and M. Urbina, "What you don't look at is what you get: Anti-saccades can reduce the midas touch-problem," in *Proceedings of the Symposium on Applied Perception in Graphics and Visualization (APGV)*, 2005.

[52] S. Hudson and G. Newell, "Probabilistic state machines: Dialog management for inputs with uncertainty," in *Proceedings of the Symposium on User Interface Software and Technology (UIST)*, 1992.

[53] S. E. Hudson, C. Harrison, B. L. Harrson, and A. LaMarca, "Whack gestures: Inexact and inattentive interaction with mobile devices," in *Proceedings of the International Conference on Tangible, Embedded and Embodied Interaction (TEI)*, 2010.

[54] T. Hueber, E.-L. Benaroya, G. Chollet, B. Denby, G. Dreyfus, and M. Stone, "Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips," *Speech Communication*, vol. 52, no. 4, pp. 288–300, Silent Speech Interfaces, April 2010.

[55] X. Huo and M. Ghovanloo, "Evaluation of a wireless wearable tongue–computer interface by individuals with high-level spinal cord injuries," *Journal of Neural Engineering*, vol. 7, no. 2, 26008, Epub 2010 March, April 2010.

[56] X. Huo, J. Wang, and M. Ghovanloo, "A magnetic wireless tongue-computer interface," in *Proceedings of the International Conference on Neural Engineering (CNE)*, 2007.

[57] T. Igarashi and J. F. Hughes, "Voice as sound: Using non-verbal voice input for interactive control," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2001.

[58] H. Istance, R. Bates, A. Hyrskykari, and S. Vickers, "Snap clutch, a moded approach to solving the Midas touch problem," in *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)*, 2008.

[59] T. Iwamoto, M. Tatezono, and H. Shinoda, "Non-contact method for producing tactile sensation using airborne ultrasound," in *Proceedings of the International Conference on Haptics: Perception, Devices and Scenarios (EuroHaptics)*, 2008.

[60] C. Jorgensen and K. Binsted, "Web browser control using EMG based sub vocal speech recognition," in *Proceedings of the Annual Hawaii International Conference on System Sciences (HICSS)*, 2005.

[61] C. Jorgensen and S. Dusan, "Speech interfaces based upon surface electromyography," *Speech Communication*, vol. 52, no. 4, pp. 354–366, April 2010.

[62] P. Ju, L. P. Kaelbling, and Y. Singer, "State-based classification of finger gestures from electromyographic signals," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2000.

[63] M. A. Just, V. L. Cherkassky, S. Aryal, and T. M. Mitchell, "A neurosemantic theory of concrete noun representation based on the underlying brain codes," *PLoS ONE*, vol. 5, no. 1, January 2010.

[64] A. Kapoor, P. Shenoy, and D. S. Tan, "Combining brain computer interfaces with vision for object categorization," in *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[65] C. Kemp, K. M. Chang, and L. Lombardi, "Category and feature identification," *Acta Psychology (Amst)*, vol. 133, no. 3, pp. 216–233, Epub 2010 January, March 2010.

[66] H. Ketabdar, A. Jahanbekam, K. Ali Yuksel, T. Hirsch, and A. Haji Abolhassani, "MagiMusic: Using embedded compass (magnetic) sensor for touch-less gesture based interaction with digital music instruments in mobile devices," in *Proceedings of the International Conference on Tangible, Embedded, and Embodied Interaction (TEI)*, 2011.

[67] J. Kim, S. Mastnik, and E. André, "EMG-based hand gesture recognition for realtime biosignal interfacing," in *Proceedings of the International Conference on Intelligent User Interfaces (IUI)*, 2008.

[68] S.-P. Kim, J. Simeral, L. Hochberg, J. P. Donoghue, and M. J. Black, "Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia," *Journal of Neural Engineering*, vol. 5, pp. 455–476, December 2008.

[69] B. Kisacanin, V. Pavlovic, and T. S. Huang, eds., *Real-Time Vision for Human–Computer Interaction*. New York: Springer-Verlag, 2005.

[70] I. Kohler, *Formation and Transformation of the Perceptual World* (originally published in 1951. Translated in 1964). New York: International Universities Press, 1964.

[71] J. Kubánek, K. J. Miller, J. G. Ojemann, J. R. Wolpaw, and G. Schalk, "Decoding flexion of individual fingers using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. v6.6, 2009.

[72] A. H. F. Lam, R. H. W. Lam, W. J. Li, M. Y. Y. Leung, and Y. Liu, "Motion sensing for robot hands using MIDS," in *Proceedings of the International Conference on Robotics and Automation (ICRA)*, 2003.

[73] G. H. Lee, "Real-time speaker adaptation for speech recognition on mobile devices," in *Proceedings of the Conference on Consumer Communications and Networking (CCNC)*, 2010.

[74] J. C. Lee and D. S. Tan, "Using a low-cost electroencephalograph for task classification in HCI research," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2006.

[75] K. Li, P. Baudisch, and K. Hinckley, "BlindSight: Eyes-free access to mobile phones," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2008.

[76] R. Lunsford, S. Oviatt, and A. M. Arthur, "Toward open-microphone engagement for multiparty interactions," in *Proceedings of the International Conference on Multimodal Interfaces (ICMI)*, 2006.

[77] R. Lunsford, S. Oviatt, and R. Coulston, "Audio-visual cues distinguishing self- from system-directed speech in younger and older adults," in *Proceedings of the International Conference on Multimodal Interfaces (ICMI)*, 2005.

[78] K. Lyons, D. Plaisted, and T. Starner, "Expert chording text entry on the twiddler one-handed keyboard," in *Proceedings of the International Symposium on Wearable Computers (ISWC)*, 2004.

[79] K. E. MacLean, "Putting haptics into the ambience," *IEEE Transactions on Haptics*, vol. 2, no. 3, pp. 123–135, July–September 2009.

[80] J. Mankoff, G. Abowd, and S. Hudson, "Providing toolkit-level support for ambiguity in recognition-based input," *Computers and Graphics*, vol. 24, no. 6, pp. 819–834, December 2000.

[81] S. Marcel, "Gestures for multi-modal interfaces: A review," Technical Report IDIAP-RR 02-34, 2002.

[82] D. J. McFarland, D. J. Krusienski, W. A. Sarnacki, and J. R. Wolpaw, "Emulation of computer mouse control with a noninvasive brain–computer interface," *Journal of Neural Engineering*, vol. 5, no. 2, pp. 101–110, 2008.

[83] D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, "Electroencephalographic (EEG) control of three-dimensional movement," *Journal of Neural Engineering*, vol. 7.3, 2010.

[84] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, and A. Kübler, "An MEG-based brain–computer interface (BCI)," *Neuroimage*, vol. 36, no. 3, pp. 581–593, Epub 2007 March, July 2007.

[85] R. Merletti and P. A. Parker, *Electromyography: Physiology, Engineering, and Noninvasive Applications.* Hoboken, New Jersey: John Wiley & Sons, 2004.

[86] P. Mistry and P. Maes, "SixthSense: A wearable gestural interface," in *Proceedings of the Asian Conference of the Special Interest Group on Computer Graphics and Interactive Techniques (SIGGRAPH ASIA)*, 2009.

[87] T. M. Mitchell, S. V. Shinkareva, A. Carlson, K. M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, "Predicting human brain activity associated with noun meanings," *Science*, vol. 320, pp. 1191–1195, 2008.

[88] C. S. Montero, J. Alexander, M. T. Marshall, and S. Subramanian, "Would you do that?: Understanding social acceptance of gestural interfaces," in *Proceedings of the International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI)*, 2010.

[89] I. Mporas, O. Kocsis, T. Ganchev, and N. Fakotakis, "Robust speech interaction in motorcycle environment," *Expert System Applications*, vol. 37, no. 3, pp. 1827–1835, March 2010.

[90] T. Nagamatsu, M. Yamamoto, and H. Sato, "MobiGaze: Development of a gaze interface for handheld mobile devices," in *Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA)*, 2010.

[91] T. Ni and P. Baudisch, "Disappearing mobile devices," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2009.

[92] E. M. Nishimura, E. D. Rapoport, P. M. Wubbels, T. H. Downs, and J. Hunter Downs, "Functional near-infrared sensing (fNIR) and environmental control applications," *Brain–Computer Interfaces, Human–Computer Interaction Series*, vol. 0, no. 2, pp. 121–132, 2010.

[93] A. V. Nurmikko, J. P. Donoghue, L. R. Hochberg, W. R. Patterson, Y.-K. Song, C. W. Bull, D. A. Borton, F. Laiwalla, S. Park, Y. Ming, and J. Aceros, "Listening to brain microcircuits for interfacing with external world — progress in wireless implantable microelectronic neuroengineering devices," *Proceedings of IEEE*, vol. 98, no. 3, pp. 375–388, 2010.

[94] I. Oakley, Y. Kim, J. Lee, and J. Ryu, "Determining the feasibility of forearm mounted vibrotactile displays," in *Proceedings of the Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS)*, 2006.

[95] S. O'Modhrain, R. Jacob, E. Gunther, and H. Ishii, "ComTouch: Design of a vibrotactile communication device," in *Proceedings of the Conference on Designing Interactive Systems Processes Practices Methods and Techniques (DIS)*, 2002.

[96] M. A. Oskoei and H. Hu, "Myoelectric control systems — a survey," *Biomedical Signal Processing and Control*, vol. 2, no. 4, pp. 275–294, ISSN 1746-8094, October 2007.

[97] A. Oulasvirta, S. Tamminen, V. Roto, and J. Kuorelahti, "Interaction in 4-second bursts: The fragmented nature of attentional resources in mobile HCI," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2005.

[98] S. Oviatt, "Taming recognition errors with a multimodal interface," *Communications of ACM*, vol. 43, no. 9, pp. 45–51, September 2000.

[99] J. Pablo Wachs, M. Kolsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," *Communications of the ACM*, vol. 54, no. 2, pp. 60–71, February 2011.

[100] T. Paek and D. M. Chickering, "Improving command and control speech recognition on mobile devices: Using predictive user models for language modeling," *User Modeling and User-Adapted Interaction*, vol. 17, no. 1–2, pp. 93–117, 2006.

[101] T. Paek, E. Horvitz, and E. Ringger, "Continuous listening for unconstrained spoken dialog," in *Proceedings of the International Conference on Spoken Language Processing (ICSLP)*, 2000.

[102] T. Paek and Y.-C. Ju, "Accommodating explicit user expressions of uncertainty in voice search or something like that," in *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech)*, 2008.

[103] T. Paek, B. Thiesson, Y.-C. Ju, and B. Lee, "Search vox: Leveraging multimodal refinement and partial knowledge for mobile voice search," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2008.

[104] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human computer interaction: A review," *IEEE Transactions on PAMI*, vol. 19, no. 7, pp. 677–695, 1997.

[105] Q. Peng and T. F. Budinger, "ZigBee-based wireless intra-oral control system for quadriplegic patients," in *Proceedings of the International Conference on Engineering in Medicine and Biology (EMBC)*, 2007.

[106] T. Pistohl, T. Ball, A. Schulze-Bonhage, A. Aertsen, and C. Mehring, "Prediction of arm movement trajectories from ECoG-recordings in humans," *Journal of Neuroscience Methods*, vol. 167, no. 1, pp. 105–114, Epub 2007 October, January 2008.

[107] T. Pylvanainen, "Accelerometer based gesture recognition using continuous HMMs," in *Proceedings of the International Conference on Pattern Recognition and Image Analysis (PRIA)*, 2005.

[108] M. B. I. Raez, M. S. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: Detection, processing, classification, and applications," *Biological-Procedures Online*, vol. 8, pp. 11–35, 2006.

[109] S. Ranganatha, A. Caria, R. Veit, T. Gaber, G. Rota, A. Kuebler, and N. Birbaumer, "fMRI brain-computer interface: A tool for neuroscientific research and treatment," *Computer Intelligence Neuroscience*, p. 25487, 2007.

[110] M. Rehm, N. Bee, and E. André, "Wave like an Egyptian: Accelerometer based gesture recognition for culture specific interactions," in *Proceedings of the British Conference on Human-Computer Interaction: People and Computers Xxii: Culture, Creativity, Interaction (HCI)*, 2008.

[111] J. Rekimoto, "GestureWrist and GesturePad: Unobtrusive wearable interaction devices," in *Proceedings of the International Symposium on Wearable Computers (ISWC)*, 2001.

[112] G. Rota, G. Handjaras, R. Sitaram, N. Birbaumer, and G. Dogil, "Reorganization of functional and effective connectivity during real-time fMRI-BCI modulation of prosody processing," *Brain Language*, [Epub ahead of print], September 2010.

[113] R. M. Rothschild, "Neuroengineering tools/applications for bidirectional interfaces, brain-computer interfaces, and neuroprosthetic implants — a review of recent progress," *Neuroengineering*, vol. 3, pp. 112, PMID: 21060801, October 2010.

[114] E. Saeedi, S. Kim, and B. A. Parviz, "Self-assembled crystalline semiconductor optoelectronics on glass and plastic," *Journal of Micromechanics and Microengineering*, vol. 18, no. 7, pp. 1–7, 2008.

[115] S. Saponas, D. Tan, D. Morris, R. Balakrishnan, J. Tuner, and J. Landay, "Enabling always-available input with muscle–computer interfaces," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2009.

[116] S. Saponas, D. Tan, D. Morris, J. Turner, and J. Landay, "Making muscle–computer interfaces more practical," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2010.

[117] S. Saponas, D. S. Tan, D. Morris, and R. Balakrishnan, "Demonstrating the feasibility of using forearm electromyography for muscle–computer interfaces," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2008.

[118] I. G. Sarason, B. R. Sarason, and G. R. Pierce, eds., *Cognitive Interference: Theories, Methods, and Findings*. Taylor & Francis, Inc., January 1996.

[119] A. Sassaroli, F. Zheng, L. M. Hirshfield, A. Girouard, E. T. Solovey, R. J. K. Jacob, and S. Fantini, "Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy," *Journal of Innovative Optical Health Sciences*, vol. 1, no. 2, pp. 227–237, 2008.

[120] G. Schalk, "Can electrocorticography (ECoG) support robust and powerful brain–computer interfaces?," *Front Neuroengineering*, vol. 24, no. 3, p. 9, June 2010.

[121] G. Schalk, K. J. Miller, N. R. Anderson, J. A. Wilson, M. D. Smyth, J. G. Ojemann, D. W. Moran, J. R. Wolpaw, and E. C. Leuthardt, "Two-dimensional movement control using electrocorticographic signals in humans," *Journal of Neural Engineering*, vol. 5, no. 1, pp. 75–84, Epub 2008 February, March 2008.

[122] R. Scherer, G. R. Müller, C. Neuper, B. Graimann, and G. Pfurtscheller, "An asynchronously controlled EEG-based virtual keyboard: Improvement of the spelling rate," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 979–984, 2004.

[123] J. Schwarz, S. Hudson, J. Mankoff, and A. Wilson, "A framework for robust and flexible handling of inputs with uncertainty," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2010.

[124] T. Scott Saponas, C. Harrison, and H. Benko, "PocketTouch: Through-fabric capacitive touch input," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2011.

[125] T. Scott Saponas, D. Kelly, B. A. Parviz, and D. S. Tan, "Optically sensing tongue gestures for computer input," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2009.

[126] S. Shenoy, "EMG in sports rehabilitation," *British Journal of Sports Medicine*, vol. 44, p. i10, 2010.

[127] M. Shilman, D. S. Tan, and P. Simard, "CueTIP: A mixed-initiative interface for correcting handwriting errors," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2006.

[128] B. Shneiderman, "The limits of speech recognition," *Communications of ACM*, vol. 43, no. 9, pp. 63–65, September 2000.

[129] J. Sodnik, S. Tomazic, C. Dicke, and M. Billinghurst, "Spatial auditory interface for an embedded communication device in a car," in *Proceedings of Advances in Computer–Human Interaction (ACHI)*, 2008.

[130] E. T. Solovey, A. Girouard, K. Chauncey, L. M. Hirshfield, A. Sassaroli, F. Zheng, S. Fantini, and R. J. K. Jacob, "Using fNIRS Brain Sensing in Realistic HCI Settings: Experiments and Guidelines," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2009.

[131] T. Starner, "The challenges of wearable computing: Part 1," *IEEE Micro*, vol. 21, no. 4, pp. 44–52, July 2001.

[132] T. Starner, "The challenges of wearable computing: Part 2," *IEEE Micro*, vol. 21, no. 4, pp. 54–67, July 2001.

[133] T. Starner, J. Auxier, D. Ashbrook, and M. Gandy, "The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring," in *Proceedings of the International Symposium on Wearable Computers (ISWC)*, 2000.

[134] T. E. Starner, "The role of speech input in wearable computing," *IEEE Pervasive Computing*, vol. 1, no. 3, pp. 89–93, July 2002.

[135] E. Stergiopoulou and N. Papamarkos, "Hand gesture recognition using a neural network shape fitting technique," *Engineering Applications in Artificial Intelligence*, vol. 22, no. 8, pp. 1141–1158, December 2009.

[136] H. Stern, J. Wachs, and Y. Edan, "A method for selection of optimal hand gesture vocabularies," in *Gesture-Based Human–Computer Interaction and Simulation,* Lecture Notes In Artificial Intelligence, vol. 5085, (M. S. Dias, S. Gibet, M. M. Wanderley, and R. Bastos, eds.), pp. 57–68, 2009.

[137] L. Struijk, "An inductive tongue computer interface for control of computers and assistive devices," *IEEE Transactions on BioMed Engineering*, vol. 53, no. 12, pp. 2594–2597, 2006.

[138] F. Tenore, A. Ramos, A. Fahmy, S. Acharya, R. Etienne-Cummings, and N. V. Thakor, "Towards the control of individual fingers of a prosthetic hand using surface EMG signals," in *Proceedings of the International Conference on Engineering in Medicine and Biology (EMBC)*, 2007.

[139] A. Toney, L. Dunne, B. H. Thomas, and S. P. Ashdown, "A shoulder pad insert vibrotactile display," in *Proceedings of the International Symposium on Wearable Computers (ISWC)*, 2003.

[140] V.-A. Tran, G. Bailly, H. Loevenbruck, and T. Toda, "Improvement to a NAM-captured whisper-to-speech system," *Speech Communication*, vol. 52, no. 4, pp. 314–326, Silent Speech Interfaces, April 2010.

[141] K. Tsukada and M. Yasumrua, "ActiveBelt: Belt-type wearable tactile display for directional navigation," in *Proceedings of the International Conference on Ubiquitous Computing (UbiComp)*, 2004.

[142] J. B. F. Van Erp, H. A. H. C. Van Veen, C. Jansen, and T. Dobbins, "Waypoint navigation with a vibrotactile waist belt," *ACM Transactions on Applications Perceptives*, vol. 2, no. 2, pp. 106–117, April 2005.

[143] Y. Vazquez-Alvarez and S. Brewster, "Investigating background \& foreground interactions using spatial audio cues," in *Proceedings of the International Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA)*, 2009.

[144] K. Vertanen and P. O. Kristensson, "Parakeet: A continuous speech recognition system for mobile touch-screen devices," in *Proceedings of the International Conference on Intelligent User Interfaces (IUI)*, 2009.

[145] S. Waldert, H. Preissl, E. Demandt, C. Braun, N. Birbaumer, A. Aertsen, and C. Mehring, "Hand movement direction decoded from MEG and EEG," *Journal of Neuroscience*, vol. 28, no. 4, pp. 1000–1008, January 2008.

[146] W. Wang, G. P. Sudre, Y. Xu, R. E. Kass, J. L. Collinger, A. D. Degenhart, A. I. Bagic, and D. J. Weber, "Decoding and cortical source localization for intended movement direction with MEG," *Journal of Neurophysiology*, vol. 104, no. 5, pp. 2451–2461, Epub 2010 August, November 2010.

[147] K. R. Wheeler, M. H. Chang, and K. H. Knuth, "Gesture based control and EMG decomposition," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 36, no. 4, 2006.

[148] A. Wilson and S. Shafer, "XWand: UI for intelligent spaces," in *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)*, 2003.

[149] A. D. Wilson, "PlayAnywhere: A compact interactive tabletop projection-vision system," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2005.

[150] A. D. Wilson and H. Benko, "Combining multiple depth cameras and projectors for interactions on, above and between surfaces," in *Proceedings of the Annual Symposium on User Interface Software and Technology (UIST)*, 2010.

[151] X. Zhang, X. Chen, W. Wang, J. Yang, V. Lantz, and K. Wang, "Hand gesture recognition and virtual game control based on 3D accelerometer and EMG sensors," in *Proceedings of the International Conference on Intelligent User Interfaces (IUI)*, 2009.

[152] J. Zhuang, W. Truccolo, C. Vargas-Irwin, and J. P. Donoghue, "Decoding 3-D reach and grasp kinematics from high-frequency local field potentials in primate primary motor cortex," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 7, pp. 1774–1784, 2010.