Sensory Substitution for Orientation and Mobility: What Progress Are We Making?

JACK LOOMIS
University of California, Santa Barbara

Sidebar 1.1 (pp. 7-10) to Chapter 1, *Perceiving to Move and Moving to Perceive: Control of Locomotion by Students with Vision Loss* by David A. Guth, John J. Rieser, and Daniel H. Ashmead (pp. 3-44). In William R. Wiener, Richard L. Welsh, and Bruce B. Blasch (Editors), *Foundations of Orientation and Mobility, Third Edition, Volume 1 (History and Theory)*, New York: AFB Press, 2010.

In the late 1960s, Paul Bach-y-Rita, Carter Collins, and their colleagues burst onto the scientific scene with tantalizing reports of a Tactile Vision Substitution System, a device that converted video images from a television camera into vibratory tactile patterns that were displayed on the user's back (Bach-y-Rita, Collins, Saunders, White, & Scadden, 1969; Collins, 1970). Their early reports indicated that users who were blind could move the camera to scan the nearby visual scene in order to recognize three-dimensional objects, intercept moving objects, and experience features of visual perspective, such as the increasing angular size of approaching objects (Bach-y-Rita, 1972; Guarniero, 1974; White, Saunders, Scadden, Bach-y-Rita, & Collins, 1970). Shortly thereafter a device was engineered that presented electrotactile displays on the abdomen; in this device electrical pulses were used instead of vibrations to create tactile sensations (Collins & Madey, 1974). Mobility studies showed that users were able to perform simple tasks under ideal conditions, such as navigating through a field of high-contrast obstacles (Collins, 1985). This work offered real hope of sensory substitution of vision by touch. In 1971, I was fortunate to land a postdoctoral position with Bach-y-Rita and Collins at the Smith-Kettlewell Institute in San Francisco and to experience firsthand the excitement and optimism of those early days. More realistic expectations gradually took hold, as it became apparent that performance with the vision-substitution system fell short of what would be needed for a true vision replacement. My experience there led to 20 years of research devoted to understanding the spatial sensing capabilities of human touch (Loomis, 1990; Loomis, Klatzky, & Lederman, 1991; Loomis & Lederman, 1986), research motivated in part by a desire to understand why performance with the vision-substitution system fell short of its early promise (for example, Apkarian-Stielau & Loomis, 1975; Loomis, 1974).

GENERAL-PURPOSE SENSORY SUBSTITUTION

Until his death in 2006, Bach-y-Rita continued working toward the goal of achieving tactile vision substitution. He eventually moved from the larger displays on the back and abdomen to much smaller displays on the upper surface of the tongue (Bach-y-Rita, Kaczmarek, Tyier, & Garcia-Lara, 1998). He and his colleagues continued using tests with simplified tasks under ideal conditions, demonstrating promising results that seemed to imply eventual success with tactile substitution of vision (Bach-y-Rita et al., 1998; Segond, Weiss, & Sampiao, 2005). In the 1990s, Bach-y-Rita was joined in the pursuit of sensory substitution by Peter Meijer (Meijer, 1992) and by Capelle, Trullemans, Arno, and Veraart (1998), who developed devices that use hearing instead of touch as the substituting perceptual modality (for an overview, see Valjamäe & Kleiner, 2006).

These tactile and auditory devices exemplify a general-purpose sensory-substitution approach, in which the stream of images from a video camera is converted into streams of tactile or auditory images, in the hope that the resulting perception will be usable for a wide range of actions involved in navigating environments and manipulating objects. A major hypothesis of the general-purpose sensory-substitution approach is that perceptual learning, as discussed later in this chapter in the section on Perceptual Learning, will occur so that people will eventually be able to process the tactile or auditory stimuli and automatically perceive relevant features of their surroundings. The approach depends on straightforward mappings of vision to touch or audition. It forgoes approaches from the field of computer vision, which preprocess the stream of input from the video camera in order to simplify it. By forgoing such preprocessing, these projects leave it up to the human user to make sense of the incoming data.

As support for the general-purpose sensory-substitution approach, advocates point to evidence from neuroscience showing that visual areas of the brain are being activated by auditory and tactile stimulation under the control of input from the video cameras (DeVolder et al., 1999; Kupers et al., 2006; for reviews, see Part 2 of Rieser, Ashmead, Ebner, & Corn, 2008). The advocates argue that eventual success is likely both because of this brain plasticity and because of the learning that comes with hundreds or thousands of hours of training

(Bach-y-Rita & Kercel, 2003; Sampiao, Maris, &, Bach-y-Rita, 2001). However, after nearly four decades of research, general-purpose systems have not been of much help with everyday O&M tasks. Because of this, I am pessimistic that general-purpose approaches to sensory substitution will ever succeed.

THE FALLACY OF THE SUCCESSFUL FIRST STEP

Research in sensory substitution has paralleled research in artificial intelligence. Both fields began with evaluations consisting of simple tasks performed under optimal environmental conditions. Like early work in artificial intelligence, the pursuit of general-purpose sensory substitution, I believe, has fallen victim to the "fallacy of the successful first step," to borrow the term of Dreyfus, Dreyfus, and Athanasiou (1986) in their critique of artificial intelligence. An example in sensory substitution has been to ask people to negotiate an obstacle course of widely separated, very high-contrast objects (Collins, 1985; Jansson, 1983). Using high-contrast objects means that critical information for performing is isolated on the tactile display, thus avoiding the need for perceptual analysis to disentangle task-critical information from extraneous background stimulation. Early success with such simplified tasks and conditions has bred optimism about eventual success in performing complex tasks in unconstrained real-world conditions, such as walking through environments cluttered with small low-contrast obstacles (for a similar critique, see Upson, 2007). Sensory-substitution devices have yet to provide for this kind of performance, and indeed, Collins (1985) noted that the success enjoyed by the Tactile Vision Substitution System with constrained navigation tasks two decades ago was not being matched by success with real-world navigation. He suggested that then-emerging technologies, such as computer vision, be incorporated to increase the likelihood of success (see also Deering, 1985). Recent research by Gonzalez-Mora, Rodriguez-Hernandez, Burunat, Martin, & Castellano (2006) is in line with Collins's suggestion.

SPECIAL-PURPOSE SENSORY SUBSTITUTION

The alternative to a general-purpose sensory-substitution approach is to design special-purpose sensory-substitution devices, each tailored for a different task. This alternative has been taken by many in the field of assistive technology for people who are blind or have low vision (Brabyn, 1985; Giudice & Legge, 2008).
The first step in the special-purpose approach is to identify what type of information and how much of that information are necessary to perform a given task. Consider the task of obstacle avoidance, in which people routinely use vision to avoid colliding with obstacles when walking through cluttered environments. One approach to determining how much information is needed for obstacle avoidance is to degrade a person's vision by limiting the field of view or spatial acuity, or both, until task performance just begins to decline. The total amount of spatial information available just as performance begins to decline is the minimum amount required for successful performance of the task. A good example of this approach is Cha, Horch, and Normann's (1992) use of low-resolution visual displays to determine the fewest number of pixels needed for travel through an environment cluttered with obstacles (see also Pelli, 1987).

The second step is to couple the critical environmental information with the substituting modality. This coupling involves two different issues: spatial bandwidth of the substituting modality and the specificity of higher-level representation. Vision, hearing, and touch each can be characterized by their spatial bandwidth, which refers to the total capacity for conveying spatial information. Given the informational requirements of a particular task (for example, navigation), it must be determined whether the spatial bandwidth of the substituting modality is adequate to handle this information. For example, no one has seriously considered using touch to substitute for vision in driving a car. Visual support of driving requires a large spatial bandwidth, both because a wide field of view is needed and because the details of critical objects, such as signs and potential obstacles, can be identified only on the basis of fine spatial detail. Regardless of how environmental information is transformed for display onto the skin, it seems unlikely that the spatial bandwidth of tactile processing will ever be adequate to allow touch to substitute for vision for driving and many other real-world tasks (Apkarian-Stielau & Loomis, 1975; Loomis, 1981, 1990; Loomis & Lederman, 1986). In contrast, as discussed above, tactile displays can be coupled to optical sensors to allow for walking to or avoiding sparsely distributed objects that stand out from the background (Collins, 1985; Jansson, 1983).

However, even if the substituting sensory modality, such as touch, has adequate spatial bandwidth to accommodate the information needed to do the task, the substituting modality may not be able to do the necessary processing—the higher-level processes of vision, hearing, and touch are highly specialized for processing the types of information that typically come through those modalities. A good illustration is the challenge of using vision to substitute for hearing. There is as yet no successful way of using vision to substitute for hearing in the reception of the acoustic speech signal. Evidence of this is the enormous challenge of visually deciphering speech spectrograms, which are graphical facsimiles of the acoustic speech signal. Even a skilled reader of spectrograms takes many times longer to decipher a speech spectrogram than to comprehend the corresponding spoken utterance. This indicates that there may be specialized brain mechanisms that operate on auditory representations of speech and that these mechanisms are not available to vision.

In the history of assistive technology for orientation and mobility, three successes stand out—for mobility, the long cane (Hoover, 1950), and for orientation, both the Talking Signs system of remote infrared audible signage (Crandall, Bentzen, Myers, & Brabyn, 2001) and GPS-based aids, such as BrailleNote GPS and Trekker (see Loomis et al., 2006; Chapters 8 and 10; and Volume 2, Chapters 11 and 14). They owe their success primarily to their provision of reliable and easy access to task-critical information while excluding other environmental information that serves only to confuse the user.

WHAT IS THE FUTURE FOR SENSORY SUBSTITUTION?

Future progress in sensory substitution for orientation and mobility is likely to come from the creation of hybrid devices that combine auditory and tactile displays with artificial intelligence preprocessing, as envisioned by Carter Collins in his prescient chapter in 1985 on mobility aids (Collins, 1985). Collins speculated that general-purpose sensory substitution involving the displav of data from a video camera or ultrasonic sensor directly to the substituting senses was unlikely to be successful. Moreover, he foresaw that O&M aids would benefit by preprocessing of the sensory data, both low-level processing, such as edge detection for simplifying the sensory stimulation sent to the display, and high-level analysis, such as for determining a clear path ahead with communication by synthetic speech. Combining computer preprocessing to eliminate extraneous environmental information with the human capacity for flexible interpretation of the incoming tactile and auditory information would seem the best chance for major advances in the development of special-purpose O&M aids.