BASIC AND APPLIED RESEARCH RELATING TO AUDITORY DISPLAYS FOR VISUALLY IMPAIRED PEOPLE

Jack M. Loomis

Department of Psychology
University of California
Santa Barbara, CA 93106
loomis@psych.ucsb.edu

ABSTRACT

This paper describes basic and applied research, conducted mostly by Roberta Klatzky, Reginald Golledge, and the author, in connection with development of a navigation system for visually impaired people. The paper begins with a treatment of the issues involved in achieving effective sensory substitution. It then discusses the importance of distance perception in spatial hearing, compares spatial language and real sound in their effectiveness in creating stable mental representations of environmental locations, and ends with a review of research on the display component of the navigation system user interface.

1. INTRODUCTION

Since the 1960's, electronic technology has given rise to a number of assistive devices for improving the lives of visually impaired and blind people. These include machines for access to optical print (like the Optacon and the Kurzweil reading machine), electronic Braille displays, obstacle avoidance devices, and, most recently, wayfinding aids like the Talking Signs® system of remote signage and GPS-based navigation systems. Leaving aside the possibility of retinal and cortical implants to replace lost visual function, assistive technology involves substitution with one of the remaining senses. For the display of spatial information, audition and touch are the viable candidates for sensory substitution. The latter comprises two submodalities, the cutaneous sense involving stimuluation of the skin surface and the kinesthetic sense, involving static and dynamic posture (e.g. of the fingers, hand, and arm). Because it is most unlikely that audition and touch together will ever provide wholesale substitution of vision, a piecemeal approach to sensory substitution is dictated. This first involves identifying various functions of everyday life, like wayfinding, reading, and interpreting graphics, that are candidates for sensory substitution. The next step is to determine what environmental information is essential for performing the function. Once the critical information has been identified, what remains is to determine if audition and touch, in some combination, might be suitable for conveying the requisite information. There are two issues here: sensory bandwidth and higher-order processing. If the sensory bandwidth of the substituting sense (or senses) is grossly inadequate, it is simply not possible to carry out the desired function. For example, the informational demands of driving a car are not likely to be met through a combination of audition and touch. However, even if the sensory bandwidth is adequate, each of the senses involves specialized cortical processing, such that one modality may not have the cortical processing that would afford successful sensory substitution. As an example, vision has yet to be exploited successfully in conveying the speech information contained within the acoustic signal to deaf people. (Lip reading is a different matter, for information about articulation is directly available in the visible signal; whereas, it is not in the acoustic signal).

The rest of this paper focuses on how best to provide a blind person with information about environmental locations beyond the range of echolocation that would be useful in navigation and environmental learning. Thus it deals with the *orientation* function of wayfinding and not at all with the *mobility* function (sensing of nearby environmental features and impediments to travel). I will mostly review basic and applied research by me and my colleagues, Roberta Klatzky (Carnegie Mellon University) and Reginald Golledge (University of California, Santa Barbara), that deals with auditory space perception and auditory display of environmental information.

1.1. Importance of distance perception in spatial hearing

Most of the scientific literature on human spatial hearing has focused on directional localization of sound sources. When people are free to turn their heads, directional localization of azimuth (compass direction) is accurate to about 1°. This is more than adequate for many of the functions served by directional hearing (such as allowing vision to rapidly localize a visible threat). The much smaller literature on auditory distance perception indicates that in natural indoor and outdoor environments with reverberation, perceived auditory distance is significantly compressed relative to source distance [1][2], even among blind people [3]. This compression is not likely a problem when a person simply wishes to walk to a continuously sounding stationary source, but it entails potentially severe consequences when a person must estimate the distance and trajectory of a moving source as, for example, when a blind person must decide whether it is safe to cross a road with fairly continuous traffic. Although it is obvious that distance misperception in such circumstances might affect a person's judgment of how much time there is to cross a road before a vehicle passes, less obvious are the effects of perceived distance errors on the shape of a moving trajectory. To the extent that perceived distance to a moving source is a nonlinear function of its distance, the shape of the source's trajectory will generally be distorted. Thus, a straight trajectory may be perceived as curved, and a curving trajectory may be perceived as straight. Similarly, the perceived shape of a configuration of stationary sources lying at different distances may be distorted.

An even less obvious perceptual distortion produced by the misperception of distance is the apparent motion of stationary sources as a person moves. In accord with Gogel's analysis of "apparent concomitant motion" of stationary visual targets that are mislocalized in distance while the person is translating [4], stationary sound sources that are mislocalized in distance may appear to move as well, as depicted in Figure 1 [5][6].

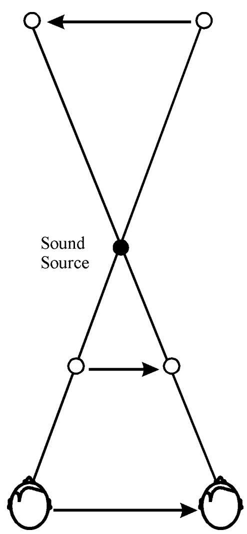

Figure 1. *Depiction of apparent concomitant motion. If the source is perceived closer than it is, it appears to move with the head; when it appears farther than it is, it appears to move in the opposite direction of the head.*

These distortions of perceived motion and perceived shape are likely to be even more severe with virtual acoustic displays, given the enormous challenge of producing displays that render distance at all realistically. This means that using a virtual acoustic display to convey motion of a source or of the shape of a configuration of stationary sources is likely to result in very significant perceptual distortion, at least until virtual acoustic displays are able to render sounds with a realism approaching that of real sounds.

1.2. Spatial updating with real sound and spatial language

One issue our research group has explored is how best to convey auditory information to people about locations in 3-D space so that they can form effective mental representations of those locations. A good way of assessing the effectiveness of the mental representations is to determine how well they can be spatially updated. Spatial updating refers to mentally keeping track of a location while moving about after the auditory (or visual or haptic) information that specified the location is no longer available. A way of measuring spatial updating performance that is independent of any errors in initially perceiving the location of the sound source is to briefly activate a sound source at a location and then require the listener to attempt to walk to its location along two separate paths on successive trials [1][3]. Successful updating is indicated by near coincidence of the two stopping points reached after traveling the two different paths. When there is perceptual error, as is typical of auditory distance perception, the two stopping points, though nearly coincident, are some distance away from the source location.

Recent research we have done has examined how well spatial language compares with sound from a loudspeaker in creating mental representations of locations, as assessed using the spatial updating procedure outlined above. On some trials, spatial language (e.g., "1 o'clock, 15 feet") was used to specify a location in a grassy field. On other trials, a loudspeaker emitting synthesized speech was placed somewhere within the field. Much to our surprise, spatial updating of locations specified by language was comparable with that of locations specified by real sound [3]. This indicates that spatial language is capable of creating stable mental representations of locations in space. A follow-up study involving the simultaneous updating of 5 locations learned just before showed that spatial language resulted in slightly poorer updating performance than real sound [7], qualifying slightly the results obtained with a single location [3].

1.3. Personal guidance system

Since 1985 our research group has conducted basic and applied research in support of our development of a navigation system for visually impaired people, that we call the Personal Guidance System (PGS). Like other GPS-based navigation systems that have been developed more recently, our system consists of three functional modules: a module for determining the position and orientation of the user, a Geographic Information System (GIS) comprising the spatial database and software providing most of the functionality, and the user interface. The original design of the PGS included use of a virtual acoustic display as the means of displaying information to the traveler. The idea was to use spatialized sound to indicate important locations in the environment, such as waypoints along the path and points of interest in the environment. As envisioned, the virtual acoustic display would present synthesized speech to the traveler by way of earphones so that the spoken names of environmental points would appear to come from their actual locations, as if emanating from loudspeakers at those locations. As part of our early efforts, we developed a virtual acoustic display using analog circuitry that proved quite effective [8].

1.4. Research on the display component of the user interface

Over the past decade, our most intense effort has focused on determining the best way of displaying information about the environment to the visually impaired user. Although the virtual acoustic display is a rich and efficient means of displaying to the user, there are several constraints dictating the need to consider other alternatives: the reluctance of some visually impaired people to wear earphones while traveling and the extra cost and complexity of using a virtual acoustic display instead of a display utilizing only synthesized speech. Consequently, our research has compared a variety of display modes, one of which utilizes virtual sound.

There are two functions of everyday life for which a display needs to be evaluated: route guidance (getting from A to B) and environmental learning (building up a mental representation of points of interest in the environment and their spatial layout). Our first evaluation of different display modes [9] compared four display modes in guiding visually impaired users over routes 71 m long and consisting of 9 segments. The different display modes continually specified the location of the next waypoint in the path until the user arrived at its location, at which time the next waypoint was activated. In the Virtual mode, a virtual acoustic display signaled the direction of the next waypoint, its number being continuously spoken by a speech synthesizer. Approximate distance information to the next waypoint was conveyed by the varying intensity of the speech as the user approached the waypoint. The other three modes indicated direction to the next waypoint using synthesized speech. The Left/Right mode specified only whether the waypoint was to the left or right of the person's body. The Bearing mode specified the number of degrees left or right to the next waypoint, and the No Compass mode was the same except that there was no electronic compass specifying the heading of the user's body; in this mode, successive GPS fixes were used to specify travel direction, which in turn was used to determine the bearing to the next waypoint. Travel performance was assessed in terms of time and distance taken to traverse the route. User ratings were also obtained. Based on these measures, the No Compass mode proved worst, the Virtual mode best, and the Bearing mode a close second [9].

Our second evaluation compared the Virtual and Bearing modes on a task of environmental learning. Users traversed a large square four times while receiving information about the locations of three off-route landmarks. After this short period of exposure, users indicated the directions to the three landmarks from different locations on the square path. This task proved very difficult, and performance was quite poor, with the two modes leading to similar performance [10]. Despite the poor performance, there is nevertheless some value in displaying such information. A person walking back and forth along the same path everyday could, over the course of several months, learn a great deal about points of interest in the environment through which they are traveling.

Research currently underway is a further evaluation of display modes in connection with route guidance. Most notable is a new mode utilizing a haptic/auditory interface, which was inspired by the hand-held Talking Signs® receiver. With the Talking Signs® system, the user localizes the direction of a infrared transmitter in the environment on the basis of the prerecorded utterances heard whenever the infrared receiver is pointed toward the transmitter. With our new interface, the user holds a lightweight rectangular block to which is attached an electronic compass. As the orientation of the user's hand varies, it is continually sensed by the computer. Whenever the pointing direction is within 10° of the bearing to the next waypoint, the computer sends a tone to the loudspeaker worn by the user. Synthesized speech indicates distance to the waypoint every 5 sec or so. This haptic/auditory interface, like the Talking Signs receiver, is very intuitive and easy to use. Its major drawback is tying up the use of one hand. One of the other modes to be evaluated in the upcoming research will use the same auditory signals as the haptic/auditory interface but the compass will instead be worn on the torso. This makes it similar to the Left/Right mode of the previous study, except that the use of tones rather than speech allows for more frequent display updates, with a consequent reduction in display lag.

1.6. Acknowledgment

The research reported here has been supported by the National Eye Institute since 1986 and by the National Institute on Disability and Rehabilitation Research since 2001 .

2. REFERENCES

[1] Loomis, J.M., Klatzky, R.L., Philbeck, J.W., and Golledge, R.G., "Assessing auditory distance using perceptually directed action", *Percept. Psychophys*. 60: 966-980.

[2] Zahorik, P., "Assessing auditory distance perception using virtual acoustics", *J. Acoust. Soc. Amer*. 111: 1832-1846.

[3] Loomis, J.M., Lippa, Y., Klatzky, R.L., and Golledge, R.G., "Spatial updating of locations specified by 3-D sound and spatial language", *J. Exp. Psychol: Learn. Mem. Cog.* 28: 335-345.

[4] Gogel, W.C., "A theory of phenomenal geometry and its applications", *Percept. Psychophys*. 48: 105-123.

[5] Loomis, J.M., "Some research issues in spatial hearing", *Proceedings of IEEE Signal Processing Workshop on Applications of Signal Processing to Audio and Acoustics*, Mohonk, NY, Oct. 15-18, 1995.

[6] Mershon, D.H., "Phenomenal geometry and the measurement of perceived auditory distance", in R.H. Gilkey and T. R. Anderson (Eds.), *Binaural and spatial hearing in real and virtual environments (pp. 257-274),* Lawrence Erlbaum Associates, Hillsdale, NJ, 1997.

[7] Klatzky, R.L., Lippa, Y., Loomis, J.M., and Golledge, R.G., "Encoding, learning, and spatial updating of multiple object locations specified by 3-D sound, spatial language, and vision", *Exp. Brain. Res*., in press.

[8] Loomis, J.M., Hebert, C., and Cicinelli, J.G., "Active localization of virtual sounds", *J. Acoust. Soc. Amer*. 88: 1757-1764.

[9] Loomis, J.M., Golledge, R.G., and Klatzky, R.L., "Navigation system for the blind: Auditory display modes and guidance", *Presence* 7: 193-203.

[10] Loomis, J.M., Golledge, R.G., and Klatzky, R.L., "GPS-based navigation systems for the visually impaired", in W. Barfield and T.Caudell (Eds.), *Fundamentals of wearable computers and augmented reality (pp. 429-446)*, Lawrence Erlbaum Associates, Mahwah, NJ, 2001.