Running head: AUDITORY NAVIGATION DISPLAYS

Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice

BRUCE N. WALKER and JEFFREY LINDSAY

Georgia Institute of Technology

Atlanta, GA, USA

Key words: Auditory display, blind navigation, sonification

Author Contact Information:
For reprint requests or other information regarding this article, please contact:
Bruce N. Walker
School of Psychology, Georgia Institute of Technology
654 Cherry Street, Atlanta, GA, USA, 30332-0170.
Phone: (404) 894-8265 Fax: (404) 894-8905
Email: bruce.walker@psych.gatech.edu

ABSTRACT

Vision loss, temporary or permanent, disrupts mobility and wayfinding. Auditory displays may assist navigation, but remain understudied.

Objectives. We examined whether spatialized non-speech beacons could guide navigation, and how sound timbre, capture radius, and practice affect performance.

Summary of background data. Existing audio navigation systems have used speech commands, leading to inefficient and cluttered displays. Non-speech sounds may provide an effective alternative or complementary approach.

Methods. 108 undergraduates navigated 3 maps, guided by 1 of 3 types of beacons (1000 Hz noise, sonar ping, or pure tone) spatialized by a virtual reality engine. Waypoint capture radius varied across subjects. Dependent measures were completion rate and efficiency.

Results. Overall navigation was very successful, with significant effects of practice and capture radius, and interactions with beacon sound. A speed-accuracy tradeoff was evident in some conditions. Waypoint hunting was exacerbated for small radius conditions.

Conclusions. Simple interfaces with non-speech beacons can effectively guide navigation. A human-scale capture radius (1.5 m) and sonar-like beacon yielded the optimal combination for safety and efficiency. Further study is required.

Implications for design/application. In addition to improving wayfinding for the blind, these findings will enable non-speech audio navigation displays for firefighters, soldiers, and others whose vision is obscured.

Navigation Performance With a Virtual Auditory Display: Effects of Beacon Sound, Capture Radius, and Practice

For a person with vision loss, the two fundamental tasks of navigating through a space and knowing what is around her can be a great challenge. At home, difficulties in navigating and learning about the environment can mean diminished mobility and increased danger. At school or college, visually impaired students may have more difficulties in simply getting from class to lab, or locating a teacher's office. They would also never know about a bench or water fountain unless they specifically asked. In the workplace, such difficulties can be an outright impediment to full participation in the corporate or urban culture. Just getting from home to work can require navigating through a mixture of indoor and outdoor spaces such as public transit stations, underground malls, city streets, and office buildings. What is a navigational challenge for sighted commuters can be nearly impossible for those with visual impairments.

There are approximately 11.4 million people with vision loss in the United States, 10% of whom have no usable vision; and by 2010 these numbers will nearly double (De l'Aune, 2002; Goodrich, 1997; National Center for Veteran Analysis and Statistics, 1994). The prevalence of blindness rises steadily with age to the extent that nearly two-thirds of people with vision loss are 65 years of age or older (De l'Aune, 2002; Goodrich, 1997). As the population of the United States ages, there will continue to be more workers with age-related visual impairments resulting from, for example, glaucoma, macular degeneration, and diabetic retinopathy. A great many of these employees can remain very productive even with diminished eyesight, so long as they are able to get to and from work, and move about the office building safely and effectively.

Spatial orientation is the major mobility problem encountered by all individuals with profound vision loss (LaGrow & Weessies, 1994; Welsh & Blasch, 1997), but is especially difficult for people whose onset of vision loss occurs later in life (Levy & Gordon, 1988; Welsh & Blasch, 1997). This includes a growing sector of the aging workforce. Wayfinding (the ability to find one's way to a destination) is dependent on the ability to remain oriented in the environment in terms of the current location and heading, and the direction of a destination. Even highly experienced blind pedestrians exhibit random movement error large enough to occasionally veer into a wall or into a parallel street when crossing an intersection (Guth & LaDuke, 1995). These problems can be compounded when the person is indoors by the lack of external orienting cues such as the sound of traffic, noise from the flow of other pedestrians, or the chirping of birds in a particular tree. While there has been a great deal of research in the area of electronic travel aids for obstacle there has not been comparable research in the development of orientation devices that keep one apprised of both location and heading (Blasch, Wiener, & Welsh, 1997). Thus, there is a critical need for navigation and orientation aids for the visually impaired.

In addition to persons with vision loss, there are whole classes of persons who have normal vision but for whom temporary smoke, fog, darkness, fire, or other environmental conditions prevent them from seeing their immediate surroundings, and can lead to disorientation and an inability to navigate from place to place. Firefighters in a smoke-filled building may not be able to locate the stairwell; military personnel in darkness or under water may not be able to reach a particular rendezvous location; police in the midst of a protest may lose orientation due to thick tear gas. Also, even when people can see, during some tasks they may be unable to use vision for navigation since it is required for another concurrent task.

It is therefore highly important to develop a system that communicates a range of information about the environment in a non-visual manner, to allow a person greater knowledge and enjoyment of, connection to, and more effective navigation through the space. Of the candidate alternative display modalities, audition is the obvious choice because of the excellent human ability to recognize and localize complex sound patterns. An appropriately developed auditory display (or *sonification*) can enhance our ability to (1) keep track of our current location and heading as we move about, (2) find our way around and through a variety of environments, (3) successfully find and follow an optimally safe walking path to our destination, and (4) be aware of salient features of the environment.

Aside from the technological challenges inherent in developing such a system, there are many questions regarding the best ways to present assistive information through an auditory interface, as well as how to design the human-system interaction. While the research in how best to design such systems has not been extensive, there have definitely been some devices developed with the goal of presenting information to support either navigation or environmental awareness. That is, in many existing navigation aides, only one of these two tasks is addressed. A system that is designed mostly for navigation might guide an individual down the street to a particular building, and then ideally to the particular room in that building she needs to reach. A system that focuses on informing the user of features of the nearby environment might address curb cuts, water fountains, or restaurants. These environmental features are not wayfinding cues, strictly speaking, but can be important for safety and quality of life reasons.

There is strong evidence, both empirical and anecdotal, of the potential benefits of auditory aids for the blind, in general (e.g., Golledge & Stimpson, 1997). This is certainly a major impetus for the development of the wearable navigation systems that use auditory displays. Basic systems such as MoBIC (Strothotte, Petrie, Johnson, & Reichert, 1995) focused on navigation but did not make use of orientation at all. Instructions on completing a pre-planned trip were communicated to the user sequentially via synthesized speech. Researchers at Sendero, LLC have more recently developed Atlas Speaks® and GPS Talk®. Atlas Speaks is a digital mapping system with synthesized voice output that can be used to plan routes of travel before leaving home. GPS Talk is a laptop computer-based system that uses a GPS mapping system to provide information about the user's location and heading, the direction of a particular destination, and limited information about things in the environment (Busboom & May, 1999). For many years, Loomis and his colleagues (e.g., Loomis, Klatzky, & Golledge, 2001) have been developing a Personal Guidance System (PGS) for the blind that uses differential GPS and compass data to guide a user along a route. Newer versions are also connected to a GIS database for information about generic places of interest such as buildings and obstacles such as phone booths. The movement directions and object information are spoken via synthesized speech. In recent versions the spoken object label is presented in 3D spatialized audio, such that the word seems to emanate from the actual location of the object. The Drishti system (Helal, Moore, & Ramachandran, 2001) has all of the features of PGS, plus a more sophisticated mapping system. It computes optimized routes based on user preference, temporal constraints (e.g. traffic congestion), and dynamic obstacles (e.g. ongoing ground work, road blockade for special events).

To date, nearly all systems have been built around synthesized speech output. The PGS (Loomis et al., 2001) is typical of the modern devices: the computer creates spoken words that sound as if they are located in the same place as the object or feature to which they refer. For example, “Doorway here” would sound as if it came from the real doorway. However, it is also important to consider non-speech audio cues, because there are several drawbacks to using exclusively speech sounds. Speech beacons are harder to localize in a virtual environment than non-speech beacons (Tran, Letowski, & Abouchacra, 2000). Users also give speech beacons low ratings for quality and acceptance (Tran et al., 2000). The speech-based interface cannot display a large amount of information, as two or more speech beacons presented simultaneously are difficult to attend to, given the limited human speech processing capacity (e.g., Mowbray, 1953; Mowbray & Gebhard, 1961). It is also difficult to use a speech-based interface for navigation and carry on a conversation at the same time (see, e.g., Wickens, 1992). Further, spoken messages in such a system are each generally more than a second long. Simpson and Marchioni-Frost (1984, cited in Stokes, Wickens, & Kite, 1988) point out that speech messages are often not understood until the whole phrase is spoken, which can make urgent messages perceived later if speech is used. In addition to the speed of processing, the length of spoken segments means that the system is often talking. For occasional spoken directions (e.g., “Turn left”), this is not a major issue. However, if the system simultaneously presenting other sounds representing the upcoming curb cut, a low hanging branch, and the location of a bus stop, the inherent inefficiency of speech can result in a cluttered listening environment (see Stokes et al., 1988 for more on this issue). While it is true that presenting a number of non-speech sounds around the user could also lead to a busy listening experience, the acoustic flexibility and brevity of non-speech sounds provides the designer with considerably more control. An immediately recognizable sound similar to trickling water could be an aesthetic and effective means of indicating the location of a fountain, without speaking “Drinking Fountain” aloud. One final concern with spoken navigation commands is that it simply takes many words to describe non-rectilinear movement: a 20° turn must be described as, “Veer to the left,” or “Turn a little bit to the left.” In our experience, simply walking toward a beacon sound is easier than translating “57 degrees” into a movement action.

Thus, while speech-based navigation sounds have been useful in some cases, there is a need to understand how to utilize non-speech sounds as well. Only recently have researchers begun to study non-speech sounds that can be used in this sort of system. Tran and colleagues (Tran et al., 2000) investigated the effect of sound characteristics on localization and subsequent navigation, the effect of real environments compared to virtual acoustic environments, and the qualitative aspects of various types of acoustic beacons. They summarize their findings by suggesting that the sounds should be wide-band and non-speech, with a proper balance between low- and high-frequency energy to make it pleasant and easy to localize. They also found that a user's rating of the quality of a sound was highly correlated with localization performance, suggesting that subjective ratings could be a useful metric for initially selecting sounds. Walker has also studied the use of non-speech sounds to convey complex information, and has looked at the attributes of the listener (e.g., Walker & Lane, 2001), the design of the sounds (Walker, 2002), and the training given to the listeners (Smith & Walker, 2002).

As part of a larger project to develop an integrated System for Wearable Audio Navigation (SWAN), we have begun to incorporate as much of the existing literature as possible into an auditory display for navigation and environmental awareness. Clearly there remain many questions about the best display design and the best interaction methods. Thus, we report here on the results of the initial studies in which the factors under investigation included (1) the different classes of sounds used as navigation beacons, (2) the effect of varying the capture radius of the navigation system (described shortly), and (3) the effects of some practice with an auditory navigation interface. These concepts are described in detail, below.

BEACONS AND CAPTURE RADIUS IN THE SWAN

The SWAN interface utilizes a repertoire of auditory icons and earcons within a specific framework to allow users to navigate successfully. The non-speech sounds in SWAN include navigation beacon sounds, object sounds, and surface transitions. These sounds are presented in a 3D audio environment, with each sound source being spatialized to seem as if it were located at the corresponding real-world location. For example, if the environmental feature the sound represents (e.g., a water fountain) is ahead and to the right of the user, the sound will appear to emanate from a location in front and to the right of the user. SWAN is able to spatialize these sounds by determining where the user is located and then placing the sounds in relation to the desired destination and the surrounding environmental features it has detected. Beacon sounds are used to accomplish the primary navigation task, while the others are used to convey knowledge about the features in the world and allow exploration of the immediate environment. We are most concerned in the present report on the beacons.

A complete path that a user might wish to travel is broken down into shorter, straight, unobstructed path segments, joined by waypoints. The beacon sounds are spatialized to emanate from the location of waypoints along the path the user is traveling. In order to travel along a preset path the user listens for the beacon of the next waypoint, and simply walks toward its apparent location. Only the immediately next waypoint is audible, to avoid any confusion that might arise from multiple concurrent waypoints. Once the user reaches the waypoint indicated by the beacon, the sound shifts to represent the location of the next waypoint, the user reorients, and sets off on the next path segment. Thus, a crucial element of the system is the ability of the user to localize the beacon sounds in the 3D audio space. Since SWAN uses generalized head-related transfer functions (HRTFs) to spatialize the sound, the more we can do to help the listener in this auditory localization task, the better. As pointed out, Tran at al. (2000) reported that specific sounds can lead to better localization, and thus better performance in this sort of task.

Each waypoint is specified by exact x, y, and z coordinates. However, the precise location of the user might never exactly "reach" the waypoint's location. Consider the following analogy: A person is supposed to start at point A, walk down the sidewalk to the corner (point B), turn left, and continue down the sidewalk to point C. There is a penny on the sidewalk at the corner, indicating the exact place to turn (the waypoint, or point B), and another penny at the destination point C. As the person walks the path she will be able to reach the waypoint, turn the corner, and complete the path successfully. However, she might never actually step right on top of the penny at either of the waypoints, despite passing pretty much right over them. A computer system might say that she failed to traverse the path correctly, since she never technically arrived at the penny-sized point. A human observer would, on the other hand, say she was definitely "close enough" to each of the points. This points to the need for a capture radius. That is, there must be a radius around the waypoint that is considered close enough, so that the next beacon sound can appear, and the user can carry on down the next path segment. If the capture radius is too small, a person might overshoot the waypoint, walk past the corner, off the sidewalk, and into the street. If the capture radius is too large, the user may be told she has reached the turning point too soon, and as a result either cut across the grass or run into the corner of a building at the intersection. Thus, to keep the person on the intended path—neither missing the marks nor turning too soon—an optimal capture radius needs to be determined.

EXPERIMENTAL VALIDATION

The first, and most important consideration is whether a fairly simple auditory interface allows users to navigate a fairly complex map successfully with no visual inputs, using non-speech beacon sounds. The next question, then, is what attributes of the display affect navigation performance, and how. We set out to test relative speed and efficiency of navigation using our system, with different non-speech beacon sounds. We also sought to study just how precisely the listeners could maneuver along a path, by varying the capture radius of the waypoints along the path. Finally, we were able to take an initial look at performance across repeated uses of the SWAN interface.

Method

Participants. A total of 108 undergraduates from the Georgia Institute of Technology participated for partial course credit (71 male, 37 female; mean age 20.2, range 18 to 30). All reported normal or corrected-to-normal vision and hearing, completed demographic surveys, and provided informed consent.

Apparatus. The auditory interface being designed is for eventual use in a complete wearable outdoor navigation system, which is under development at Georgia Tech. In order to study a variety of aspects of the SWAN interface we have developed a virtual reality-based prototyping environment. This allows us to implement and rapidly evaluate our sounds, menus, and interaction devices, in a safe and controlled lab environment before testing with the full SWAN system. Our VR environment was constructed using the Simple Virtual Environments (SVE) software package developed by the College of Computing at the Georgia Institute of Technology (GVU Virtual Environments Group, 1997). SVE was run on a Dell Optiplex PC running at 1.7Mhz, with 528 MB of RAM. The beacon sounds were played through closed ear headphones. To change direction participants rotated on the spot where they were standing. They used two buttons on a joystick to control forward and backward movement in the VR environments (they did not actually walk forward). Their orientation within the environment was tracked by an Intersense InertiaCube 2 head-mounted tracking cube attached to the headphones. We have ensured that the auditory interaction, including physically orienting to 3D audio sounds, is identical in the VR environment and full SWAN systems. Other than the movement method, initial testing with participants has revealed that users of both systems do not report any major differences in the experience.

Each participant was asked to navigate three different paths (or <u>maps</u>) with the auditory navigation interface. The environment in which these maps were located was essentially a large empty (virtual) room with four walls. In addition to the starting point, Map 1 had five waypoints, and Maps 2 and 3 each had 10 waypoints. The three maps differed simply in the layout of the waypoints and in overall length. The <u>scheduled map length</u> is the sum of the lengths of the shortest-distance path segments. That is, if a person moved precisely from waypoint to waypoint to waypoint in a map, she would travel a distance equal to the scheduled length. The scheduled length of the maps was 100.0, 283.5, and 287.6 units, respectively. In the virtual environment, one unit of distance is approximately equal to one meter of real distance. The SVE software logged the participant's location (in terms of X, Y, and Z coordinates), head orientation (angular pitch, yaw, and roll), and the waypoint she was currently moving towards, every 2 ms.

The participants were divided into three groups, with each group being guided through the maps by a different navigation beacon sound. The beacon sounds for all three groups were 1 s long, with a center frequency of 1000 Hz and equal loudness. The sounds differed greatly in timbre, however. The first sound beacon was a burst of broadband noise centered on 1000 Hz. This particular sound had the broadest spectrum of the three. The second beacon was a pure sine wave with a frequency of 1000 Hz. The third beacon sound was a sonar pulse, similar to the sound that Tran el al. (2000) found to be one of the best sounds for use as a navigation beacon.

Each participant navigated using the same sound throughout their three maps. At the start of a map the beacon sound played in an on-off pattern, where the sound was on for 1 s and off for 1 s of silence (50% duty cycle). As the listener moved closer to the next waypoint the silence was shortened to effectively make the beacon tempo faster. Hence, increasing proximity to the waypoint was mapped to increasing tempo, which is consistent with our findings for population stereotypes or preferred mappings between proximity and tempo (Walker, in preparation).

Within each beacon-sound group, one third of the participants had a small (0.5 m) capture radius, one third had a medium (1.5 m) capture radius, and the final third had a large (15 m) capture radius. Thus, both the beacon sound and capture radius were constant throughout the three maps for a given participant.

Procedure. Each participant was randomly assigned to use one of the three beacon sound types, with a total of 36 participants using each sound. The experimenter explained the salient aspects of the beacons—namely that the sound is spatialized to indicate the relative direction of the next waypoint and that tempo is mapped to distance from that waypoint—and discussed potential front-back confusions that can sometimes occur with artificially spatialized sounds and non-individualized HRTFs. Participants then learned how to use a combination of body rotations and joystick button presses in order to move through the environment. Once the study began, participants moved through the three maps one after the other, with a brief rest between maps. The map order was the same for all participants. Following completion of the third map, the experimenter debriefed and thanked the participant.

Results

We first considered the global question of whether participants would be able to complete the navigation tasks using only non-speech auditory cues. Figure 1 presents the movement traces of all participants in Map 2, for each combination of capture radius and beacon sound (results for the first and third maps are very similar). The straight dark solid line between the waypoints represents the scheduled path, and the other lines in each panel represent the actual paths traveled by the participants. The first result to note is the relatively successful navigation through the map by nearly all participants. Not only could participants complete the maps, they picked up the task very quickly and with little instruction. Nevertheless, it is important to note that in some cases there are significant departures from the scheduled path. Most often these result from a participant walking just past a waypoint and not realizing it for some time because the beacon sound is mislocalized as coming from the front instead of from the rear (i.e., an overshoot exacerbated by front-back confusion). This navigation error occurs most often with the smallest capture radius. In the left column of panels of Figure 1 (the smallest capture radius) several of the waypoints have a “star-like” pattern of movement traces around them. This is the result of a participant overshooting the waypoint, turning around and heading back towards it, then overshooting again. This hunting behavior does not appear nearly as often for the medium capture radius, and is very rare for the largest capture radius. It is interesting to note that the very erratic movement by one or two participants with the sonar beacon and large capture radius seems a general failure to navigate, and is not limited to hunting for a waypoint.

The second result to note is the difference between the general movement patterns in the different capture radius conditions. In the smallest capture radius condition participants stick very close to the scheduled path, and pass precisely over the waypoints enroute. There is a sort of “pinch point” at each waypoint that is very small for the small radius (naturally). In the larger two capture radius conditions the pinch point is more relaxed, and if a person strays off the scheduled path, he or she need not come exactly back to the path in order to carry on—the capture radius allows some flexibility (or “slop”, depending on one’s perspective) in the path. For the medium radius the participants seem to move off the path in some cases, but still come back to the waypoint.

Finally, in the largest capture radius condition, participants often never even reach the actual waypoint. They come close enough for the capture radius to be satisfied, but their overall path is actually quite different, geometrically, from the scheduled path. The turning angles are often considerably more or less acute than the angles in the path they were supposed to travel. Certainly the severity of this depends on the context and the reasons for which the person is traversing the path. While different combinations of beacon sounds and capture radius conditions tend to lead to somewhat different types of performance, for practical purposes, the medium capture radius has a compromise between relatively little overshooting and hunting, and relatively close passage by the waypoints. At this point we turned to a more quantitative analysis of the rate of completion and the path length efficiency in the various conditions.

We analyzed the data using a three-way mixed factors multivariate analysis of variance (MANOVA). The two between-subjects factors were the beacon sound used and the capture radius. The within-subjects factor was map number. The dependent measures being recorded for each map were the participant’s overall completion rate (map length divided by time to complete the map) and their navigation efficiency (total distance traveled divided by scheduled map length). Since participants typically veered off the shortest path at least sometimes, and in some cases overshot the waypoints, the actual distance the participants traveled was usually (but not always) longer than the scheduled map length. The “extra” distance they traveled can be considered as wasted time and effort, and we reasoned that comparing the distance the person was supposed to travel to the distance they actually followed would be a useful metric of the movement efficiency afforded by the different beacon sounds and capture radii. This efficiency metric can also be viewed as an indicator of how effective the map might be in guiding a visually impaired person along a specific path (e.g., along a sidewalk). With this in mind, deviation from the path could potentially be very dangerous if there were environmental hazards just off of the path (e.g., roads, a ditch, etc.). Thus, a priori we assumed an optimal efficiency score to be 100 percent, which would indicate that for the most part the participant had stayed very close to the scheduled map route. In the multivariate analyses we used Wilks' Lambda to determine F values, and throughout all analyses we set an alpha level of .05.

The results of the MANOVA on the combined dependent variables revealed a significant effect of which map was being traversed, $F(4, 96) = 103.03$, $p < .001$, Wilks' Lambda = .19, and a significant effect of the capture radius being employed, $F(4, 196) = 63.67$, $p < .001$, Wilks' Lambda = .19. There was also a significant multivariate interaction of map number and capture radius, $F(8, 192) = 8.95$, $p < .001$, Wilks' Lambda = .53. These significant multivariate effects led us to seek further clarification in the results for the two dependent variables considered separately. Before contemplating the univariate results we checked the data for outliers and for violations of the assumptions of normality, linearity, outliers, homogeneity and multicollinearity, with no serious violations noted. We did apply the Greenhouse-Geisser correction to the degrees of freedom in significance tests for effects on efficiency, in order to correct for violations of sphericity (Mauchley's $W = .822$, $p < .001$). This was unnecessary in the case of rate ($W = .957$, $p > .05$).

The overall mean completion rate and efficiency for each map are shown in Figure 2. There was a significant increase in rate as participants completed Maps 1 through Map 3, with mean rates of 0.76, 1.23, and 1.29 distance units per second, respectively, $F(2, 198) = 219.96$, p

$< .001$. There was also an increase in efficiency for subsequent maps, with mean efficiencies of 74.8, 92.7, and 99.0 percent, respectively, $F(1.7, 189.8) = 99.99$, $p < .001$. Both of these results reflect an overall main effect of practice with the system.

There was also a main effect of capture radius on both rate, $F(2, 99) = 29.07$, $p < .001$, and efficiency, $F(2, 99) = 24.16$, $p < .001$. These results are presented in Figure 3. In the case of rate (Figure 3, left panel), overall the largest capture radius yielded the fastest completion rate (1.44 units/s), the medium capture radius led to the slowest rate (0.75), and the smallest capture radius led to an intermediate rate (1.08). In the case of efficiency, however, the results are quite different (see Figure 3, right panel). The largest capture radius led to a moderate efficiency (88.1%), while the medium capture radius led to the greatest efficiency (105.8%), and the smallest radius led to the lowest efficiency (72.5%). Note that efficiency can be greater than 100% since the implementation of a capture radius makes it possible to traverse a path that is actually shorter than the scheduled map length. Taken together, these two results for rate and efficiency are effectively a speed-accuracy tradeoff. For example, in the case of the medium capture radius the participants were slow but very efficient. Participants using the large capture radius were fast but somewhat inefficient. That is, they spent less time orienting themselves to the beacon sounds, and subsequently traversed a longer path than necessary. However, the large capture radius was very “forgiving”, and as a result they were still able to complete the maps quickly. The importance of these various strategies will be discussed shortly.

In addition, these two main effects were moderated by an interaction of map and capture radius, but only for rate, $F(4, 198) = 9.82$, $p < .001$, and not for efficiency, $F(3.4, 168.0) = 10.3$, $p > .05$. For the sake of comparison, results for the interaction of map and capture radius are shown for both rate and efficiency in Figure 4. For rate (Figure 4, left panel), participants using the largest capture radius started with the highest rate, and then improved the most. The medium capture radius led to the lowest rate in Map 1, and led to the smallest improvement over the course of the experiment. The smallest capture radius led to an intermediate rate on Map 1, and an intermediate level of improvement with practice. It is the difference in improvement for the three capture radius groups that leads to the significant interaction. In the case of efficiency (Figure 4, right panel), the total amount of improvement with practice was not different for the three groups.

There were also significant interactions involving the beacon sounds heard by the participants. There was a marginal main effect of beacon sound for efficiency, $\underline{F}(2, 99) = 2.54$, $\underline{p} < .08$, with an ordering in terms of efficiency of the noise beacon (highest, at 94%), then the pure tone (90%), and finally the sonar ping (83%). The order of performance was the same for rate, namely noise (1.18 units/s), pure tone (1.09), and sonar (1.00), though the effect did not reach conventional levels of statistical significance, $\underline{F}(2, 99) = 1.95$, $\underline{p} < .15$. There was, however, a significant interaction of beacon sound and map for rate only, $\underline{F}(4, 198) = 3.10$, $\underline{p} = .017$. As seen in Figure 5, the noise beacon led to the fastest rates, as well as to the greatest increase in rate with practice. The pure tone and sonar ping beacons led to slower rates and to less improvement with practice. Finally, there was a significant interaction of beacon sound and capture radius for efficiency only, $\underline{F}(4, 99) = 2.62$, $\underline{p} = .04$, which is shown in Figure 6. The interaction comes from the fact that there were differences in efficiency for the three beacon sounds in the large capture radius condition, but no differences among beacon types at the medium and small capture radius conditions. This seems to indicate that the choice of beacon sound affects efficiency, but only for the very large capture radius.

DISCUSSION

There are several important ideas to be drawn from the results presented here. The first and most important is that the non-speech auditory interface can definitely be used for successful navigation. Participants were able to follow the paths in the virtual environment using only the spatialized beacon sounds. Their ability to do so is well illustrated by the traces through Map 2, shown in Figure 1. With absolutely no other cues but the navigation sounds, even in the least effective cases most participants strayed relatively little from the path designated by the beacons, and all were able to complete the maps. The same pattern is exhibited in the results from Maps 1 and 3, as well. This successful performance amongst almost all individuals is evidence that the interface leads to successful navigation through the virtual environment. In the physical world the additional navigation cues already present in the acoustic ecology, as well as the additional sensory information from the ground, a cane, wind on the face, and so on, will only make the informational environment richer, leading to even better performance. This is important since the likelihood of simultaneous conversation, use of a radio or mobile telephone, or other speech communication points to the need for a non-speech navigation system.

In the few cases where a participant’s path did deviate significantly from the beacon path, it was most often due to overshooting a beacon by passing just outside its capture radius. Once that happened the participant might have experienced front back confusion and did not turn around to find the beacon because it still sounded as if it was ahead of them. This can lead to a dramatic departure from the planned route, so it must not be dismissed. In debriefing participants it seems clear that there are listeners who just do not seem to “get” the interface, and never really navigate as well as the rest. It may be important to isolate what leads to such confusion with the navigation cues. However, we should be clear that these instances are quite rare, in our experience. Most people pick up the task immediately, show good performance from the start, and steady improvements with practice. We have noted that the overshoot likelihood is exacerbated by the smallest capture radius. Thus, given that occasionally participants will just miss the target waypoint, we have considered a number of ways to make passing by the waypoint more salient, in addition to not using the smallest capture radius, of course. Studies of a variety of ways to increase the salience of waypoint passing, and the employment of front-back disambiguation methods are currently underway and will be reported elsewhere.

Next, the effect of capture radius on performance was found to be more practically significant than that of beacon sound. That is, capture radius seems to be a more important design consideration than beacon sound in the construction of an auditory navigation interface. Tran et al. (2000) investigated the effectiveness of various types of beacon sounds, but did not consider other potentially important factors. The results of the present study provide evidence that while beacon sounds should certainly be considered and evaluated, there are other critical aspects of an interface to factor in.

The third important point is that practice is significant. Participants' performance based on both rate and efficiency improved dramatically across trials. Practice also interacts significantly with capture radius and with beacon sound, with regard to rate. Performance continued to improve, even in Map 3, suggesting that the limit to gains in performance based on practice for this task has likely not yet been reached. Further study is needed to understand more clearly what type of relationship exists between extensive continued practice and navigation performance as well as what the limits to improvement with practice may be.

Given the significant interactions, it is important to note that each of the factors must be considered together. An example of this can be seen in Figure 6, where the sonar beacon showed a much worse performance in terms of rate when combined with a large capture radius, but shows no significant difference in performance for the medium or small capture radius. It seems then practical to consider each aspect of the interface in relation to the others as changes to one could have potentially important implications for others. This also highlights the difference between theoretical and practical considerations. Since our interface is designed to be used for movement along a set path, safety (or remaining on the path) is a paramount concern. There is an obvious speed-accuracy tradeoff occurring between rate and efficiency for different capture radii. Given that fact, and given our primary concern for accuracy, we would first look at the capture radius that led to the best efficiency, and then consider other factors that may affect rate (as well as different beacon sounds) as the situation permits. In a real world application it does not matter if a person using SWAN to navigate down the sidewalk does not move quite as quickly, so long as he or she can manage to remain on the sidewalk throughout the path. Thus, as is often the case, a true human-centered approach must be taken in order to avoid "optimizing" the system at the expense of the user. All things considered, we would conclude that a capture radius of approximately 1.5 m should be optimal for auditory navigation. With that choice, the sonar ping and pure tone led to slightly faster and more efficient paths than the noise beacon (which only performed better in the large capture radius condition). In that case, the sonar ping is likely preferred, as it is a more complex tone, less subject to masking by environmental noises.

The development of a non-speech auditory interface of this type remains a work in progress. The results presented here are clearly applicable to movement in a virtual environment. As the result of pilot studies in an actual movement situation (i.e., not in the virtual reality environment), and our own experience with the outdoor SWAN, it is evident that the localization of the beacons and the interaction with the system remains similar, so the overall navigation is robust. We do, however, note some differences in the actual movement style that the users employ. For example, we have noticed that outdoors users tend to start out walking a little bit more slowly with the system, but this effect quickly diminishes with continued usage and increased confidence. With the important results of the present investigation, we will now be able to extend usage in the full SWAN system to continue to validate the veracity of the virtual environment. Since such an interface is novel for all users, in addition to the general effects of interface design elements we are beginning to study the effectiveness of different training methods on performance. This includes an evaluation of the basic learnability of different interface sounds and the most effective types of training. Further, it is not clear whether there are individual differences in the perception, understanding, and learning of auditory displays (speech or non-speech), nor how one might predict performance with such a system. Also, to our knowledge, none of the audio-based navigation systems to date involves context- or task-dependent adjustments to the information that is presented. The needs of the listener, within her present acoustical and functional environment, must be factored in so the interface can adapt appropriately. For example, if a user is on target to a waypoint 30 m down a straight hall, with no obstacles in the way, then the system could (or perhaps should) stay relatively quiet and let the person use the mobility skills she already has. Approaching the target, the system would gracefully chime in again. A related issue is communicating to the listener the degree of certainty about location, orientation, and items in the surroundings. Knowing that there is some uncertainty in the location (perhaps due to relying solely on GPS) may lead the user to adjust attention and other movement techniques. In the present tests of the interface, the exact location is known, and the listener need not rely on any other sensory input for guidance. Finally, it remains to be studied how effectively a user can navigate with an auditory wayfinding system, while at the same time completing other cognitive tasks such as decision making and planning.

This multitask proficiency will be an important measure of the success of any system aimed at assisting in navigation.

In summary, we have shown the effectiveness of non-speech auditory beacons in guiding a listener along a path, and have presented rate and path efficiency as useful metrics that need to be considered together when designing and evaluating auditory navigation interfaces. The actual beacon sounds employed in the interface are important to consider, but so, too, is the way the user interacts with the sounds, including the implementation of a capture radius about the waypoints on a path. Specifics of the user's task need to be considered as well. Applications such as our System for Wearable Audio Navigation (SWAN) need to place efficiency and accuracy ahead of speed, in order to maximize safety from the outset. Ongoing and planned work will determine the best ways to then improve on speed and multitasking for users.

REFERENCES

Blasch, B., Wiener, W., & Welsh, R. (1997). *Foundations of Orientation and Mobility* (2nd ed.). New York: American Foundation for the Blind.

Busboom, M., & May, M. (1999). Mobile navigation for the blind. *Proceedings of the International Conference on Wearable Computing* (pp., Vienna, Austria.

De l'Aune, W. (2002). *Legal Blindness and Visual Impairment in the Veteran Population 1990-2025*. Decatur, GA: VA Rehabilitation R&D Center.

Golledge, R., & Stimpson, R. J. (1997). Spatial Cognition, Cognitive Mapping, and Cognitive Maps. In *Spatial Behavior: A geographic perspective* (pp. 224-266). New York: Guilford Press.

Goodrich, G. L. (1997). *Growth in a Shrinking Population: 1995-2010*. Palo Alto, CA: Palo Alto Health Care System.

Guth, D., & LaDuke, R. (1995). Veering by blind pedestrians: Individual differences and their implications for instruction. *Journal of Visual Impairment & Blindness, 89*(1), 28-37.

GVU Virtual Environments Group. (1997). *SVE Toolkit*. Retrieved November 24, 2000, from http://www.cc.gatech.edu/gvu/virtual/SVE

Helal, A., Moore, S., & Ramachandran, B. (2001, October 7-9). Drishti: An Integrated Navigation System for Visually Impaired and Disabled. *Proceedings of the Fifth International Symposium on Wearable Computers (ISWC'01)* (pp., Zurich.

LaGrow, S., & Weessies, M. (1994). *Orientation and mobility: Techniques for independence*: Royal New Zealand Foundation for the Blind, Dunmore Press.

Levy, S. B., & Gordon, A. R. (1988). Age-related vision loss: Functional implications and assistive technologies. *International Journal of Technology and Aging, 1*(2), 16-125.

Loomis, J., Klatzky, R., & Golledge, R. (2001). Navigating without vision: basic and applied research. *Optometry and Vision Science, 78*(5), 282-289.

Mowbray, G. H. (1953). Simultaneous vision and audition: The comprehension of prose passages with varying levels of difficulty. *Journal of Experimental Psychology, 46*, 365-372.

Mowbray, G. H., & Gebhard, J. W. (1961). Man's senses as informational channels. In H. W. Sinaiko (Ed.), *Human factors in the design and use of control systems* (pp. 115-149). New York: Dover.

National Center for Veteran Analysis and Statistics. (1994). *National Survey of Veterans*: Assistant Secretary for Policy and Planning, Department of Veterans Affairs, Gov't Printing Office.

Smith, D. R., & Walker, B. N. (2002, 02-05 July). Tick-marks, axes, and labels: The effects of adding context to auditory graphs. *Proceedings of the 8th International Conference on Auditory Display* (pp. 362-367), Kyoto, Japan.

Stokes, A., Wickens, C. D., & Kite, K. (1988). *Display technology : human factors concepts.* Warrendale, PA: Society of Automotive Engineers.

Strothotte, T., Petrie, H., Johnson, V., & Reichert, L. (1995, 26 - 28 April). MoBIC: an aid to increase the independent mobility of blind and elderly travelers. *Proceedings of the 2nd TIDE Congress* (pp., Paris.

Tran, T. V., Letowski, T., & Abouchacra, K. S. (2000). Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics, 43*(6), 807-827.

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. *Journal of Experimental Psychology: Applied, 8*(4), 211-221.

Walker, B. N. (in preparation). *Stability of magnitude estimation for auditory data representations.*Unpublished manuscript.

Walker, B. N., & Lane, D. M. (2001). Psychophysical scaling of sonification mappings: A comparision of visually impaired and sighted listeners. *Proceedings of the 7th International Conference on Auditory Display* (pp. 90-94), Espoo, Finland.

Welsh, R., & Blasch, B. (1997). *Foundations of orientation and mobility* (2nd ed.). New York: American Foundation for the Blind.

Wickens, C. D. (1992). *Engineering psychology and human performance* (2nd ed.). New York, NY: HarperCollins Publishers.

FIGURE CAPTIONS

Figure 1. Movement traces for all participants in each combination of beacon sound and capture radius, while moving through Map 2. Participants were able to complete the course with little practice and instruction. Some overshoots and bouncing are noted, and this differed across conditions of capture radius and beacon sounds. See text for further explanation.

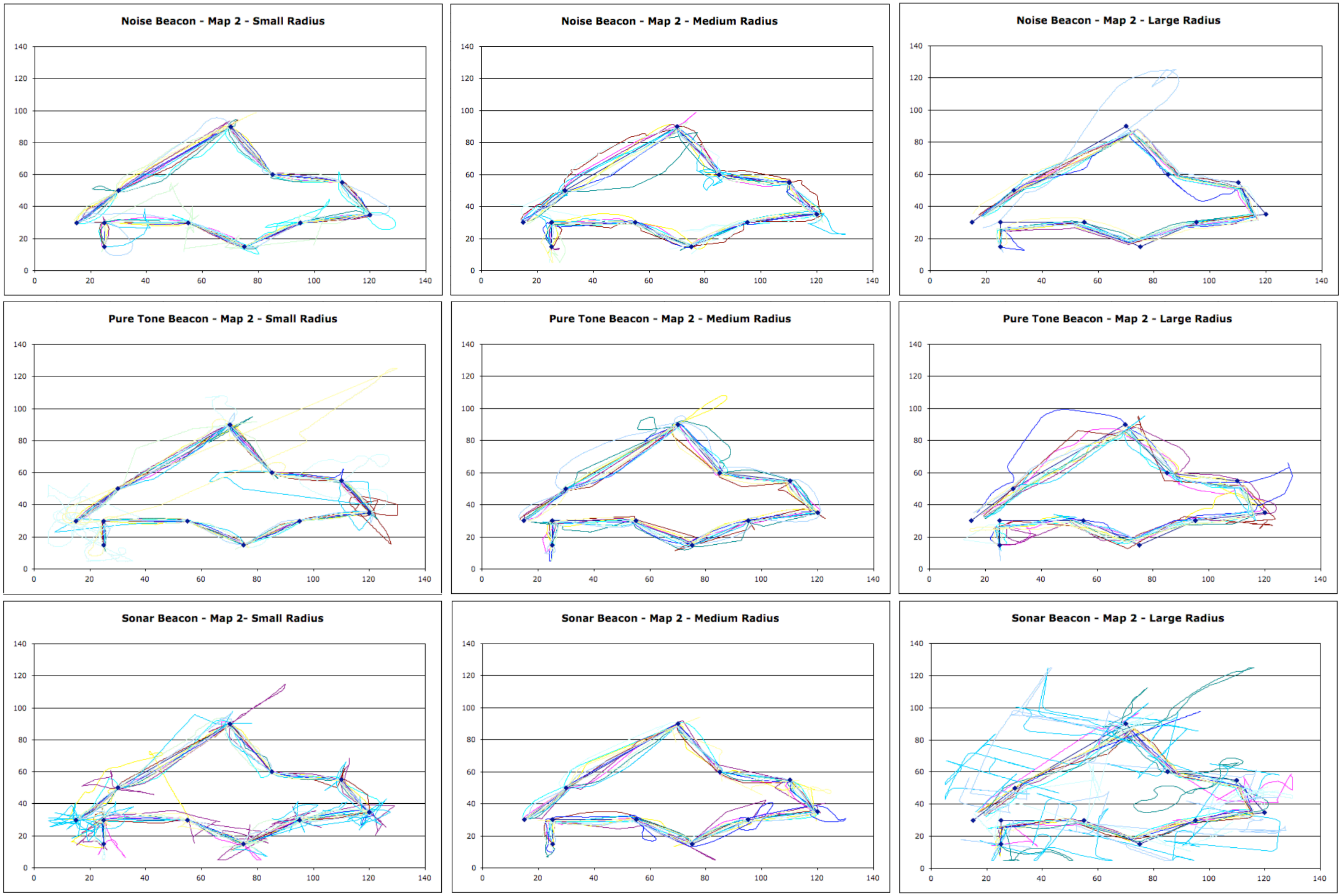
Figure 2. Main effect of map on completion rate (left panel) and efficiency (right panel). The significant effects indicate an overall improvement in performance from map to map, namely a practice effect.

Figure 3. Effect of capture radius on completion rate and efficiency. The main effects indicate speed (rate) – accuracy (efficiency) tradeoff in performance.

Figure 4. Interaction of map and capture radius on completion rate and efficiency. The interaction reached statistical significance only for rate (left panel), and not for efficiency (right panel). The effect for rate indicates a differential practice effect for the three capture radius groups.

Figure 5. Interaction of beacon sound and map on completion rate. The noise beacon led to the highest completion rate as well as to the largest gain in rate across maps. The sonar beacon led to the slowest rates and least improvement, while the pure tone beacon led to intermediate results for rate.

Figure 6. Interaction of beacon sound and capture radius on efficiency. The three beacons led to different patterns of performance across the three capture radius conditions.



Noise Beacon - Map 2 - Small Radius

Noise Beacon - Map 2 - Medium Radius

Noise Beacon - Map 2 - Large Radius

Pure Tone Beacon - Map 2 - Small Radius

Pure Tone Beacon - Map 2 - Medium Radius

Pure Tone Beacon - Map 2 - Large Radius

Sonar Beacon - Map 2- Small Radius

Sonar Beacon - Map 2 - Medium Radius

Sonar Beacon - Map 2 - Large Radius

Figure 1.

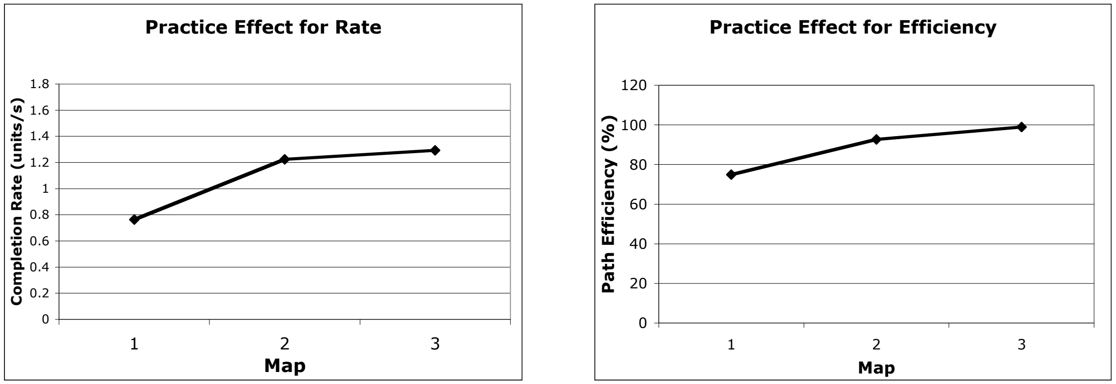

Practice Effect for Rate
Completion Rate (units/s)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
2
3
Map
Practice Effect for Efficiency
Path Efficiency (%)
120
100
80
60
40
20
0
1
2
3
Map

Figure 2.

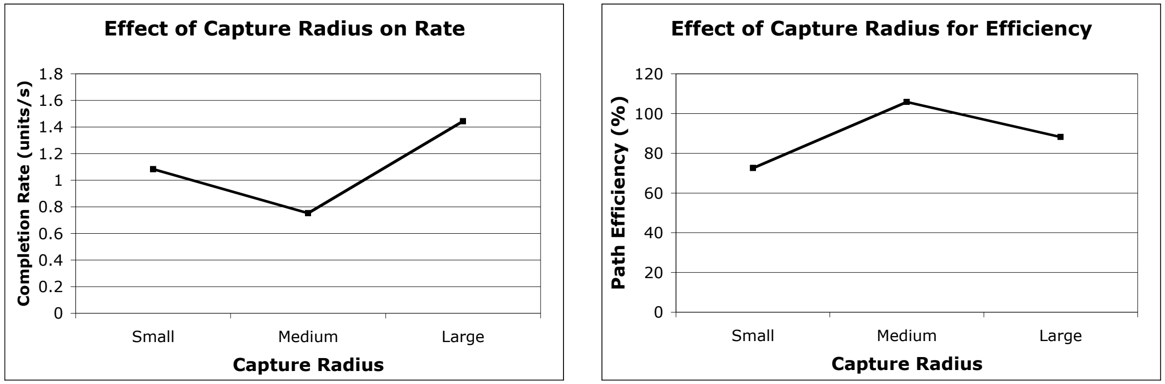

Effect of Capture Radius on Rate
Completion Rate (units/s)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Small
Medium
Large
Capture Radius
Effect of Capture Radius for Efficiency
Path Efficiency (%)
120
100
80
60
40
20
0
Small
Medium
Large
Capture Radius

Figure 3.

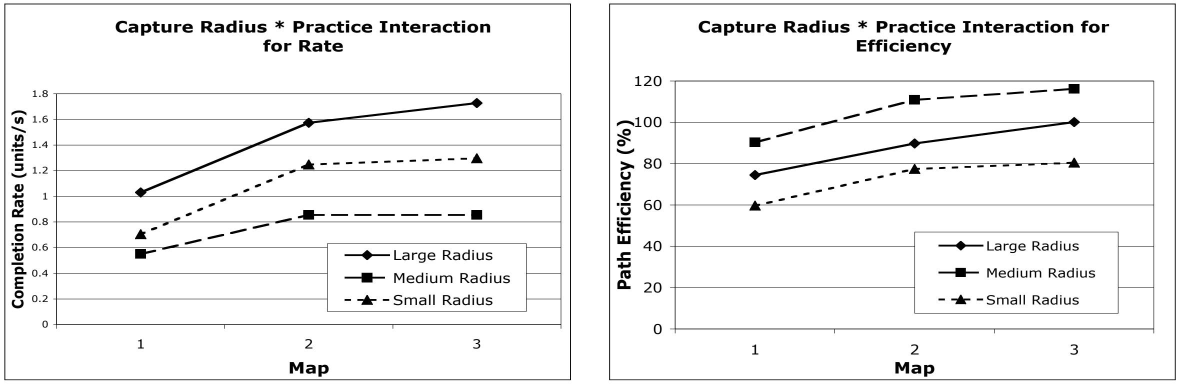

Capture Radius * Practice Interaction for Rate
Completion Rate (units/s)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
2
3
Map
Large Radius
Medium Radius
Small Radius
Capture Radius * Practice Interaction for Efficiency
Path Efficiency (%)
120
100
80
60
40
20
0
1
2
3
Map
Large Radius
Medium Radius
Small Radius

Figure 4.

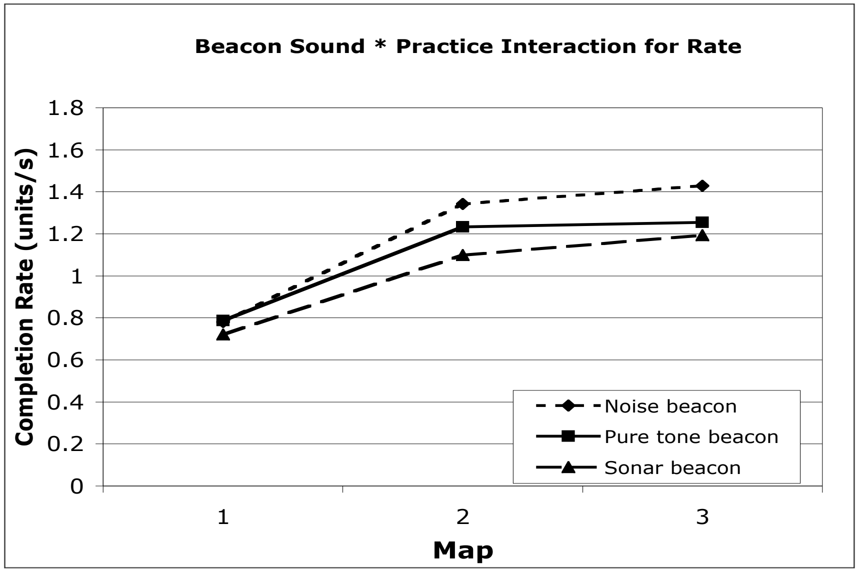

Beacon Sound * Practice Interaction for Rate
Completion Rate (units/s)
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1
2
3
Map
Noise beacon
Pure tone beacon
Sonar beacon

Figure 5.

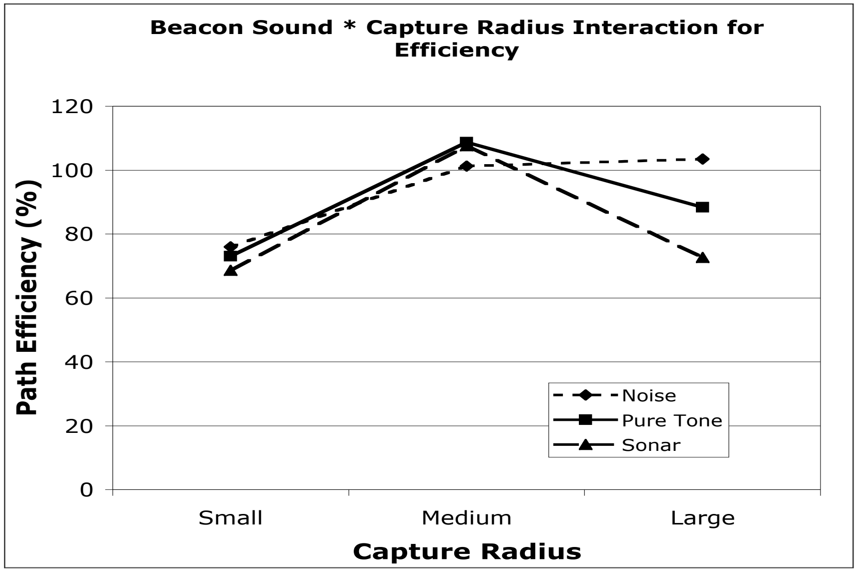

Beacon Sound * Capture Radius Interaction for Efficiency
Path Efficiency (%)
120
100
80
60
40
20
0
Small
Medium
Large
Capture Radius
Noise
Pure Tone
Sonar

Figure 6.

BIOGRAPHIES OF THE AUTHORS

Bruce N. Walker is an assistant professor in the School of Psychology and the College of Computing at the Georgia Institute of Technology. He received his Ph.D. in Psychology in 2001 from Rice University.

Jeffrey Lindsay is a graduate student in Engineering Psychology at the Georgia Institute of Technology. He received his B.S. in Psychology in 2002 from Georgia Tech.