KnowWhere™: An Audio/Spatial Interface for Blind People

Dr. Myron W. Krueger
Artificial Reality Corporation
Box 786
Vernon, CT 06066
860-871-1375
860-871-9445, 7738 FAX
myronkru@aol.com

Dr. Deborah Gilden
Smith-Kettlewell Eye Research Institute
2232 Webster Street
San Francisco, CA 94115
415-561-1665
415-561-1610 fax
debby@skivs.ski.org

ABSTRACT

The KnowWhere™ System was developed to present geographic information to blind people. The user's hand rests upon an illuminated surface which has been covered with a tactile grid and is viewed by a ceiling-mounted video camera. The video image is analyzed by specialized processors and the location of the user's fingertip on the light table is determined. An invisible virtual map is defined on the desk surface and the feature that the user is currently pointing to is signalled by a sound which serves to identify the feature or kind of feature that has been "touched."

BACKGROUND

The KnowWhere™ system grows out of a long abandoned dissertation project implemented in 1969.[1][2][3] That project, called Z-Meister, was intended to teach assembly language programming. It was built around an interpreter that was instrumented so that every program was accompanied by graphics and by auditory patterns generated on a Moog synthesizer, thus giving the user a sense of what their program was doing as it ran. Since then, we have used sounds as a debugging technique on a routine but unsystematic basis.

The sensification of programs expanded to physically experiencing them and led to Dr. Krueger’s interactive environment and artificial reality concepts. From the beginning of the VIDEOPLACE system, sound has been used to substitute for touch when a participant’s video image came into contact with a graphic object.[1,2,3] The idea of using audio-touch to present information to blind people dates back to the beginning of his work, but did not resonate with anyone in the disabilities community until 1994 when Dr. Krueger and Dr. Gilden met and collaborated on this NIH sponsored SBIR Phase I project.[4]

RELATED WORK

Others have used sound in the human-computer interface to illustrate system functions [5][6][7] and for debugging [8][9]. In addition, sound has long been used to try to make information accessible to blind people starting with the audio oscilloscope.[10] More recently, both researchers and commercial entities has sought to make graphical user interfaces to blind people.[11][12]

KnowWhere™

In the KnowWhere™ system, the data is not sonified so much as sound is used to substitute for the tactile sensation of touching a virtual object. The user’s hands resting upon the illuminated surface of the VIDEODESK are viewed by an overhead camera. An invisible virtual map is defined on that surface and as users move their index fingers around it, they receive audio feedback which alerts them to the feature or kind of feature they have touched.

The sound may be a recording of a human speaking the name, an abbreviated form of the name, a text-to-speech generated name, or any one of a variety of sound codes. If users want to examine a particular area in more detail, it can be enlarged to the size of the entire desk. In this mode, the user can explore much smaller features, noticing bays, lakes, estuaries, and straight versus irregular boundaries.

The VIDEODESK is a direct pointing device so blind users know where they are pointing and can train their proprioceptive sense so they can find features faster as they gain experience with the system. In contrast, the mouse is a relative pointing device, and therefore is almost impossible for blind individuals to use.

MAP INFORMATION PRESENTED

Basic geographic information was presented about the United States and the countries of South America, Europe, and Africa. The capitals and other features were also available.

GESTURE SAMPLING AND SONIFICATION

KnowWhere™ seeks to provide a false sense of touch by combining the proprioceptive sense of the location of the finger on the desktop with the sound that we provide as the user touches a feature on the virtual map. This illusion requires the perfect synchronization of perceived audio events with expected tactile sensations.

If there is a lag between the time a user touches a virtual feature on the desktop and the time when the user hears the sound that is triggered by the contact, the feature will be perceived to be at the location where the finger is when the sound is heard, not at its true location. The faster the user's finger is moving at the time, the greater the discrepancy.

Thus, the sample rate determines the size of feature that can be detected as the user's hand moves at a given speed, because a narrow feature that lies between two consecutive samples will be invisible to the system. The user can compensate by slowing down the movement of his or her finger.

Currently, the VIDEODESK interface delivers 30 samples per second which would be fast enough for the user to connect finger position with map location if there were no other sources of delay. However, the sound responses themselves have duration. While a sound response is being played, the system ignores other features being touched.

We tried aborting the sound from the first feature and starting the next when it was contacted. However, the resulting concatenation of word fragments creates sounds which are sometimes bizarre. For instance, the passage through Montana, Idaho, Washington, and Oregon, might result in the system uttering "Mont-Ida-Wa-Oregon."

Representational sounds like ocean waves or city traffic are not immediately recognizable. They can take several seconds to recognize, even when they are selected by name. Also, many features such as mountains, rivers, lakes, etc do not have unambiguous sound associated with them. For this reason, we abandoned representational sounds.

There was a similar problem with the instrumental sounds and sound effects which we used in some of the sonification modes. A percussive sound like a bell takes a full second to play. Even clicks and pops are .25 seconds long.

For spoken feedback, the problem is even more severe. It takes 1.5 seconds to say "Massachusetts" or "Mediterranean Sea". While there are methods for marginally speeding up both recorded and generated speech, the resulting utterances are still longer than we would like.

We experimented with audio abbreviations of the names. The first syllables such as "Col" for Colorado, "Cal" for California, "Kan" for Kansas, "Tenn" for Tennessee, and "Ken" for Kentucky were not easy to distinguish from each other.

To our surprise, it was often faster and more understandable to say two syllables even if the result did not correspond to any normal concept of abbreviation. For instance, "rado" stood for Colorado, "sota" stood for Minnesota, "sippi" stood for Mississippi, and so on. These audio abbreviations were quite easy to learn and use when the names being replaced were familiar.

When a single state was expanded to fill the whole screen, there was no problem of ambiguity. A single very brief click was used to indicate when the finger was inside, a metallic ding was used when it was on the boundary, and silence was heard when it was outside. With this technique, users could distinguish straight, irregular, and smoothly curved boundaries. They were quite able to examine detailed features such as estuaries, the tiny gap between upper and lower Michigan, the barrier islands off the coast of North Carolina, and the deeply indented coastline of Rhode Island. Indeed, blind users could observe that sighted individuals seldom notice because they seldom look at more detailed maps even though they are available to them.

BLIND SUBJECTS

Five subjects tested the KnowWhere™ system. All were middle-aged. There were two congenitally blind men, two congenitally blind women, and one adventitiously blind man. One of the women is a guide dog user and all the others use canes. Three of the subjects were quite skeptical of the idea beforehand. Two of these were completely converted within the first seconds of use. The third remained resistant throughout, although her performance improved markedly.

All of the subjects were able to use the interface to find the states or countries whose locations they had just learned as well as those with which they were already familiar. We were impressed by how good a sense of absolute location the subjects were able to acquire. All were able to identify puzzle pieces of the features that they had just explored. Two could even produce quite acceptable drawings of the features they had examined.

COMPARISON WITH TACTILE MATERIALS

While tactile maps would seem to be superior to an audio-spatial presentation, they are difficult to create, difficult to store, and they are seldom produced at the size or level of quality needed to demonstrate their superiority. While the variety of printed maps for the sighted is virtually limitless, relatively few raised-line maps are produced and most blind people have little access to them. For the half of the blind population that cannot read Braille, the labels on these materials are useless unless assistance is provided.

What most severely limits the resolution of tactile maps is the fact that the Braille annotation must compete for space with graphic features. For instance, an 11” by 14” map of Africa does not have enough room to completely label most of the countries. Many are labelled with only one or two letters. “BF” does not suggest Burkina Fasso to a user who has never heard of the place. Finally, tactile maps offer no means of zooming in to see more detail.

ADVANTAGES OF KnowWhere™

KnowWhere™ provides a flexible interface to maps for blind people. Its audio/spatial interface could also be used to provide access to other kinds of spatial information. Because the components are potentially small and do not require any non-electronic special fabrication, KnowWhere™ lies in the mainstream of technological development. Video cameras already can be implemented on a chip using the same fabrication techniques that are used to create microprocessors. Thus, the camera and computer-vision algorithms could be integrated on the same chip in the near future. While software implementations are possible currently, these are much too slow for this purpose. The time lag they introduce makes the audio-spatial sense much less immediate. However, if processors continue to get faster or if ever more complex operating systems did not confiscate speed improvements as fast as they are delivered, a completely software solution would be possible. At that point, assuming that a video camera will be bundled with the computer as has already begun to occur, the KnowWhere™ technology will be free.

FUTURE WORK

The current sound feedback only provides information about the point on the map at the tip of the finger. This tunnel vision means that the user is not getting as much shape information as through tactile materials which can be investigated with several fingers simultaneously. The VIDEODESK processing hardware upon which the KnowWhere™ system rests can provide instantaneous positions for up to 40 fingers without slowing down the application at all. Using two fingers in concert, the users will be able to measure features. They will also be able to use a finger on one hand to menu while they point to features on the map with the other. This would allow them to quickly switch among multiple overlays displayed over the same geographic area.

At the same time, we will investigate sonification techniques which will provide the user a richer experience of the feature he has touched. We have identified a number of sonfication techniques that we can use to provide the user with a view of the local curvature of the shape. Thus, while our initial effort makes a modest point about how to use sound to substitute for touch in a practical context, the anticipated developments will address interesting issues in sonification itself.

REFERENCES

1. Krueger, M. Computer Controlled Responsive Environments, Dissertation, University of Wisconsin, 1974, p. 178.
2. Krueger, M. Artificial Reality, Addison-Wesley, 1983.
3. Krueger, M. Artificial Reality II, Addison-Wesley, 1991.
4. Krueger, M. “KnowWare™: Virtual Reality Maps for Blind People,” SBIR Phase I Final Report, NIH Grant #1 R43 EY11075-01, 1996.
5. Gaver, W. “Using and Creating Auditory Icons,” ICAD 92.
6. Buxton, W., “Using Our Ears: An Introduction to the Use of Nonspeech Audio Cues” in Extracting Meaning from Complex Data: Processing, Display, Interaction, edited by E.J. Farrel, Vol. 1259, SPIE 1990, pp. 124-127.
7. Cohen, J. “Monitoring Background Activities.” In Auditory Display: Sonification, Audification, and Auditory Interfaces.” edited by G. Kramer. Addison-Wesley, 1994.
8. Francioni, J., L. Albright, and J. Jackson. “Debugging Parallel Programs Using Sound.” In Proceedings of the ACM/ONR Workshop on Parallel and Distributed Debugging, Addison-Wesley, 1991,

9. DiGiano, C., and R. Baecker. “Program Auralization: Sound Enhancements to the Programming Environment.” In Proceedings of the Graphics Interface ‘92. pp. 44-52.
10. TEKWEEK, “Scope displayed for visually impaired.” November 7, 1980
11. Mynatt, E. & G. Weber. “Nonvisual Presentation of Graphical User Interfaces: Contrasting Two Approaches.” CHI 94.
12. Dufresne, A., O. Martial, C. Ramstein, and P. Mabilleau. “Sound, Space, and Metaphor: Multimodal Access to Windows for Blind Users.” ICAD ‘96.