

Audio Enhanced 3D Interfaces for Visually Impaired Users

Stephen W. Mereu, Rick Kazman

Department of Computer Science

University of Waterloo

Waterloo, ON, Canada, N2L 3G1

+1 519 888 4567 x4870

rnkazman@cgl.uwaterloo.ca

ABSTRACT

Three dimensional computer applications such as CAD packages are often difficult to use because of inadequate depth feedback to the user. It has, however, been shown that audio feedback can help improve a user's sense of depth perception. This paper describes an experiment which evaluates the use of three different audio environments in a 3D task undertaken by visually impaired users. The three audio environments map tonal, musical, and orchestral sounds to an (x, y, z) position in a 3D environment. In each environment the user's task is to locate a target in three dimensions as accurately and quickly as possible. This experiment has three important results: that audio feedback improves performance in 3D applications for all users; that visually impaired users can use 3D applications with the accuracy of sighted users; and that visually impaired users can attain greater target accuracy than sighted users in a sound-only environment.

Keywords

User Interface, Auditory Interface, Disability Access, 3D Interface

INTRODUCTION

Audio feedback has proven itself to be an effective addition to human-computer interfaces in numerous studies. Many computer applications use sounds to present information to the user that would otherwise be difficult to convey through other means. Audio is also useful when the user's attention needs to be focussed away from a visual display. For example, Gaver, Smith and O'Shea demonstrated the utility of audio feedback with their Arkola cola factory simulator. In this experiment users "listened to" processes that were not visible in their view of the factory [4]. The bottle capper machine, for instance, normally made a rhythmic sound. When something went wrong with the capper, the rhythm changed thus alerting the user.

Along with monitoring background events, complex foreground events can also be aided using audio feedback. Di-

Giano and Baecker, for example, added sound to a programming environment [1]. By attaching a different sound to each routine or block of code, semantic errors, like infinite loops, were easily detected. Jackson and Francioni also used sound in a programming environment, but they focused on conveying the state of executing parallel programs to the user [6]. Although audio can help users monitor processes and events, it can also be used to listen to raw data. Hayward demonstrated this by attaching sounds to seismic signals [5]. By doing so, significant seismic events were then quickly determined by listening to changes in the data. The possibilities for listening to data are virtually endless. Known applications range from listening to the stock market, company annual sales, computer load, or even bodily functions like a heart beat and respiratory rate [3].

One use of audio feedback that has proven to have significant ramifications is improving the access of visually impaired users to computers. This is an increasingly important area of study as user interfaces are becoming more graphical and less text-based and thus hindering the visually impaired user (who previously managed reasonably well in text-based environments by using speech synthesis to "read" the contents of the screen). Some work has already been performed in the area of improving access to graphical environments for visually disabled users. Mansur, Blattner and Joy's Sound-Graphs showed how a sound could be mapped to (x, y) data so that visually impaired users could perceive the relationship between x and y. They did this by scanning along the x axis and adjusting the pitch of the sound according to the y value [7].

When given a choice between a text based application and a windows based one, visually impaired users prefer the text based product since it is easier for them to use. Edwards sought to help visually impaired users in using a windowing system [2]. He succeeded by attaching an identifying tone to each window including the edge of the screen. As the user moved the cursor around the screen, each entered window would play its identifying tone. The user would then know which window was the active window.

Sound can help visually impaired users function in a windowing interface—a 2D graphical environment. Can it also help them in a 3D environment: a CAD package, or a vir-

Permission to make digital/hard copies of all or part of this material for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copyright is by permission of the ACM, Inc. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires specific permission and/or fee.

CHI 96 Vancouver, BC Canada

© 1996 ACM 0-89791-777-4/96/04..\$3.50

tual environment? Such environments are difficult for normally sighted users, principally because they try to convey depth information in a 2D environment, typically through the use of orthogonal views. Sound has proven to be effective in improving the performance of users of such 3D environments [8].

If sound can be similarly used for visually impaired users, it would break down a substantial barrier, thus allowing visually impaired users access to 3D applications. The research of this paper seeks to answer this question, and gives reason to believe that this barrier can be surmounted.

PREVIOUS WORK

Before studying the effects of sound in a 3D computer interface with visually impaired users, we studied its effect with sighted users. This initial study served four purposes: a) to determine if sound could be used as an effective *depth cue* to improve depth perception in 3D interfaces, b) to compare the differences between various types of sound cues, c) to note how user performance changes when all visual aids are removed thus forcing users to rely solely on sound cues, d) to study the effect that learning has on user performance while using a sound based computer environment. [8]

The first three questions were studied using twenty sighted subjects. Subjects were presented with a randomly rotated “blobby” object that had a target dot randomly placed on its surface. Figure 1 shows an example blobby object. The user's task was to move the cursor to the target—shown as a white spot on the object's surface in Figure 1—in 3D space. Cursor movement was controlled using a standard mouse for the x and y directions and the up/down arrow keys for the z dimension. When the cursor was at the desired location, the user depressed the space bar signaling that the task was complete. Subject performance was measured by automatically recording both their target accuracy (distance from the target) and task time.

Subjects were tested on four types of environments: one that gave no audio feedback, a tonal environment that provided audio feedback by altering a simple sine wave tone, a musical environment that altered a piece of music being played for the user, and an orchestral environment that used an orchestral arrangement to play music.

During a task trial with each of the audio environments, the subject would continually “hear” the cursor's location. Upon depressing the left mouse button, however, the user would hear only the sound of the target location. **By pressing and releasing the mouse button consecutively an audible comparison is made so that the subject can determine the similarity between the two sounds, and hence the relationship between the cursor and the target.**

The tonal sound environment attached a simple sine wave tone to each location. As the cursor moved in 3D, the sound would change in three of its *sound dimensions*, such

as pitch or volume. Pretest experiments found that the best mapping of spatial dimensions to sound dimensions was x (left/right) to balance (left/right), y (up/down) to pitch (high/low) and z (far/close) to volume (quiet/loud) [8].

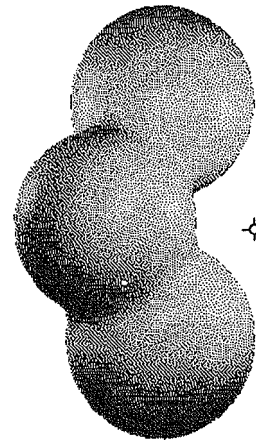


Figure 1: “Blobby” Object with Target Location

The “blobby” object's irregular shape and lack of hard edges minimize the visual depth cues provided to the user. Both the object's orientation and the location of the white target are randomly chosen per trial to reduce learning effects. Users moved the crosshair cursor in 3-space to locate the target dot.

In the musical environment a very similar approach was taken, but instead of modifying a simple tone, a randomly selected musical piece was played. This piece was modified according to the cursor's location. As with the tonal environment, pretests were carried out which determined the best sound dimension mappings. These were, as with the tonal environment, x to balance, y to pitch and z to volume.

The orchestral environment extended the idea of playing music by arranging a set of instruments into a standard configuration and using this to locate a sound in 2D. The initial motivation for this mapping was the observation that tonal and musical sounds are good for giving users relative position information but, excepting those rare users with perfect pitch, they are inadequate for giving absolute position information. An orchestra, on the other hand, is a fixed configuration and so, after a small amount of training, when a user hears a trumpet play, they know what location this represents. Absolute position information was thought to be valuable for visually impaired users.

In the orchestral environment the x-z horizontal plane was partitioned into eight instrument sections as shown in Figure 2. Each section was assigned a different instrument. Instruments were selected based on pretest results that determined the eight most identifiable instruments from the 128 available in the MIDI patch set.

Using this environment, the music played corresponding to a given cursor location would be dominated by the instrument in the cursor's orchestra section. For example, if the cursor was located in the 1/8th of the x-z plane that was

occupied by the violin section, the violins would play the melody the loudest while the other sections would play a background part. Thus the orchestra metaphor provided the x and z components of the cursor's 3D location. The third spatial component, y, was mapped to oscillation. This mapping was selected based on pretest results.

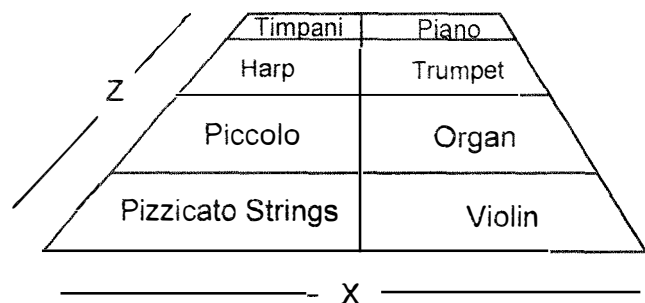


Figure 2: Orchestra Arrangement

The orchestral environment uses an orchestra consisting of eight instrument sections laid out on the x-z horizontal plane.

User performance in these three audio environments, plus a visual-only environment, was then studied.

RESULTS OF THE INITIAL STUDY

Twenty paid normally sighted subjects performed 25 trials on each of the four environments. All subjects performed the experiments in a quiet room using a Pentium¹ PC running Windows 3.1² equipped with a Gravis Ultrasound³ card, a pair of headphones and the experiment application software. Subjects were instructed on each sound environment before beginning and were told that both speed and accuracy were equally important.

The order of exposure to the four environments was randomized. The set of trials for each environment was divided into 5 *learning* trials, 15 *visual* trials and 5 "*blind*" trials. After they had completed the learning and visual trials, the subjects performed what were termed the "*blind*" trials. These were exactly like the normal visual trials except the screen was blanked and the subjects had to rely entirely on audio feedback. In addition to the recorded measures of target error and task time, four subjective ratings were also measured per sound environment: subjective performance, preference, usability, and marketability.

Results showed that all three audible sound environments (tonal, musical, and orchestral) reduced target errors over the no sound environment by 78.9%, 61.1% and 32.4% respectively. Task time, however, suffered and increased over the no sound environment by 123.4%, 178.4% and 215.1% respectively.

The no visual "*blind*" tasks caused an increase in target error over the visual cases in only one of the three audio

environment: orchestral sound. Normally sighted users were thus able to accurately locate the target in a "*blind*" environment relying only on tonal or musical feedback. **This suggested the possibility of using the sound environments in a sound-only application when a sighted user needed to focus elsewhere, or in an application for visually impaired users.**

Overall, subjects performed the best, in terms of speed and accuracy, using the tonal environment. However, they found tonal sound to be annoying and preferred to use the musical environment. Subject differences such as gender, area of study, musical ability and previous graphics experience made no significant difference to the results. The order that the environments were tested also caused no significant difference.

The results of this experiment prompted a second question: how would the results change when subjects were exposed to the sound environments over an extended period? Another experiment was therefore run with seven paid subjects. Subjects performed the same experiment on a Monday, Wednesday, and Friday of a single week. Results showed that target error remain the same as before, but task time was reduced significantly over the three days. Given that target error decreased but task time increased over the no sound environment in the initial study, this result is encouraging. It shows that, with a small amount of practice, the additional time required to attend to the audio cues can be mitigated.

AN EXPERIMENT WITH VISUALLY IMPAIRED USERS

The three sound environments (tonal, musical, and orchestral) proved to be effective position cues and, more importantly, depth cues, for normally sighted users. As a result we thought it likely that visually impaired users could also benefit from these audio environments. In a subsequent study eight paid visually impaired subjects (3 low vision and 5 blind) were used to test this hypothesis. The results of this study are the main topic of this paper.

The experimental design was similar to the previous study described above, having 25 trials per sound environment (tonal, musical, and orchestral) this time using 10 as learning and 15 as experiment trials. Since the no sound environment is a visual only environment, it was not tested with blind subjects and was used only as a learning tool for the low-vision subjects.

Results: Visually Impaired Subjects

The three sound environments were compared using one-way ANOVA tests with repeated measures on the sound environment. A significant difference in target error was found across the sound environments with $F(2,14)=87.12$, $p=0.0001$.

Figure 3 shows the difference in these target errors. The error level in the tonal environment was approximately

1. Pentium is a registered trademark of Intel.

2. Windows is a registered trademark of Microsoft Corp.

3. Gravis Ultrasound is a registered trademark of Advanced Gravis Computer Technology Ltd.

half that of the musical environment and almost a fifth of the orchestral environment.

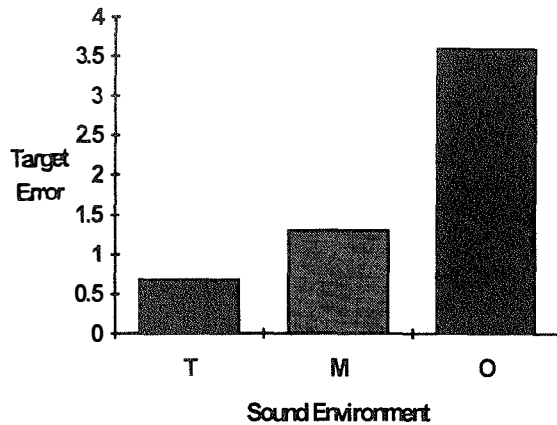


Figure 3: Sound Environment vs. Target Error

The above graph shows that errors were the least when subjects were using the tonal sound environment.

No significant difference was found in the task time across the three sound environments, as shown in Figure 4, with $F(2,14)=0.75$, $p=0.4924$.

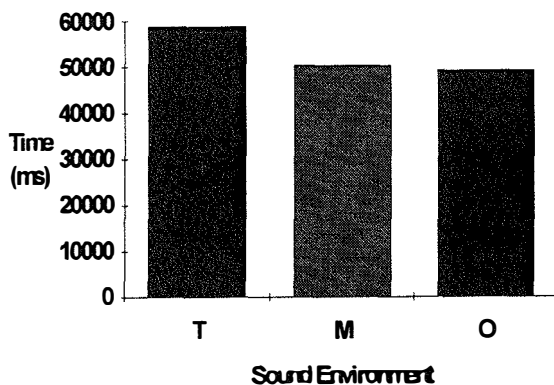


Figure 4: Sound Environment vs. Time

The time required to reach the target were approximately the same for each sound environment and showed no significant difference.

The visually impaired users rated the tonal sound environment higher than both the musical and orchestral environments with respect to subjective performance, preference, usability, and marketability, as shown in Figure 5. These differences, however, did not reach the 0.05 level of significance, likely due to the small number of subjects.

Sighted vs. Visually Impaired Subjects

Since the normally sighted subjects had already been tested on the three sound interfaces, it was of interest to see if there was any significant differences between their performance and preferences and the performance and preferences of the visually impaired subjects. As mentioned previously, the normally sighted subjects were tested on the sound environments both with and without visual cues,

representing a *combined audio and visual* application and a *sound-only* application respectively. Both of these tests were compared to the responses of the visually impaired subjects.

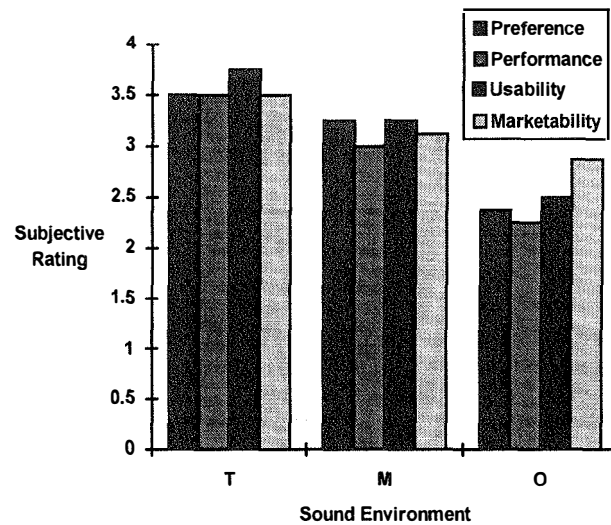


Figure 5: Sound Environment vs. Subjective Ratings

In this figure larger subjective results are better. The tonal sound environment was rated as the best sound environment over the musical and orchestral sound environments on all four of the subjective measures by visually impaired users.

Sighted Subjects with Audio and Visual Cues

The target error and task times of the sighted and visually impaired subjects showed a significant interaction with the sound environment. Figures 6 and 7 show the interactions of target error and task time respectively. It is not surprising to see that the sighted subjects had better accuracy and time throughout since they had the use of both audio and visual cues.

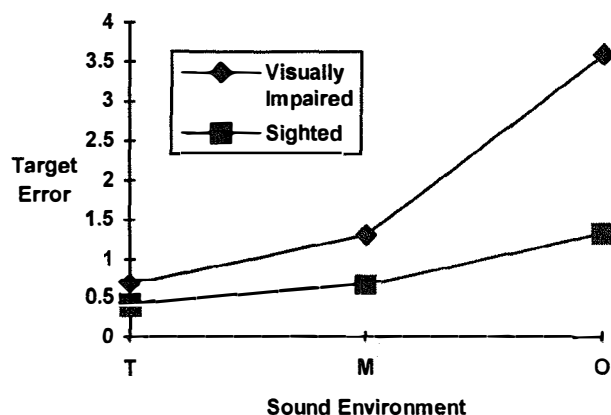


Figure 6: Interaction of Visual Ability and Sound Environment with Respect to Target Error

This graph shows the interaction of the subject's visual ability and the sound environment. Sighted users had the use of both audio and visual cues, whereas visually impaired users had only audio cues.

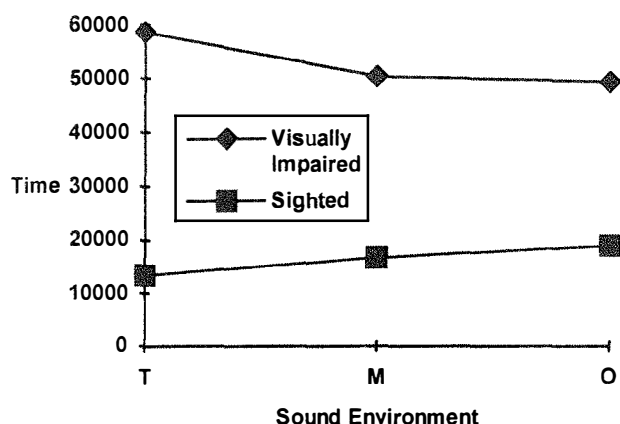


Figure 7: Interaction of Visual Ability and Sound Environment with Respect to Time—With Visuals

As in Figure 6, the sighted subjects had the use of both audio and visual cues whereas the visually impaired subjects used only the audio cues. Sighted subjects were able to complete the task significantly faster than visually impaired subjects.

What is interesting is that although the visually impaired subjects took, on average, 4.3 times as long using the tonal environment, their accuracy (0.68) was comparable to the sighted user with the musical environment (0.67). This means that the tonal sound environment by itself is giving the visually impaired users as much information as the combination of visual and tonal cues gives a normally sighted user. It is not surprising that the times for visually impaired users were longer: the normally sighted users located the (x, y) position of the cursor very quickly, by simply placing the crosshair over the white target spot, and spent most of their time using the sound cues to locate the proper depth (z).

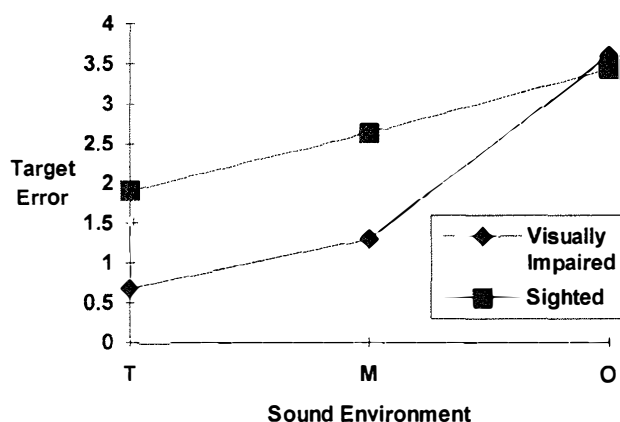


Figure 8: Interaction of Visual Ability and Sound Environment with Respect to Target Error

This graph is the same as Figure 6 except the sighted subjects could not use visual cues. Both the sighted and visually impaired subjects received only audio cues. The visually impaired subjects had target errors significantly *less* than the sighted subjects in both the tonal and musical environments.

The visually impaired users also performed as well with the musical environment (1.30) as the sighted user did with the orchestral environment (1.33). Least Significant Difference (LSD) T tests were performed on each sound environment comparing the responses of the two visual abilities and showed a significant difference in all cases at $\alpha=0.05$.

Sighted Subjects with Sound-only

When the sighted subjects had only audio feedback as did the visually impaired subjects, their performance degraded as seen in Figures 8 and 9. Target error for sighted subjects in the tonal and musical environments were now significantly *worse* than the visually impaired subjects' when compared using an LSD test at $\alpha=0.05$. The orchestral sound environment's target error level showed no significant difference between the two visual abilities.

In addition to more target errors, the sighted users also took longer with this sound-only scenario than when they had both visual and audio cues. LSD tests at $\alpha=0.05$ showed no significant difference between the time responses of the two visual abilities per sound environment.

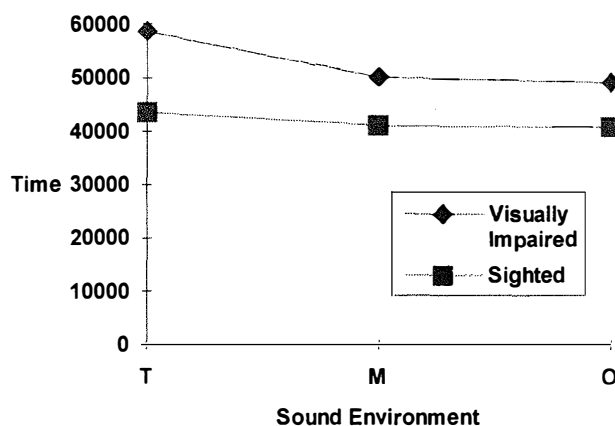


Figure 9: Interaction of Visual Ability and Sound Environment With Respect to Time—No Visuals

When sighted users were without visual aids their task times were significantly higher than in Figure 7. Compared to visually impaired users' task times, no significant difference was found.

Subjective Differences

The subjective measures (performance, preference, usability, and marketability) were all compared using ANOVA tests to note any differences between the sighted and visually impaired subjects. For subjective performance, usability, and marketability no significant interaction with the three sound environments was detected with $F(2,52)=0.48$, $p=0.62$; $F(2,52)=0.74$, $p=0.70$; $F(2,52)=0.48$, $p=0.62$ respectively. There was also no significant difference in the three subjective rating due to visual ability with $F(1,26)=1.05$, $p=0.32$; $F(1,26)=0.40$, $p=0.53$; and $F(1,26)=3.04$, $p=0.09$ respectively.

Subjective preference, however, did show a significant interaction ($F(2,52)=3.54$, $p=0.04$). Figure 10 shows this interaction with visually impaired subjects preferring the tonal environment and the sighted subjects preferring the musically-based environments.

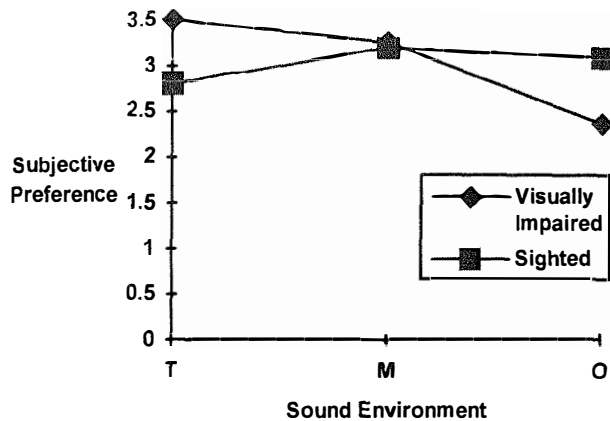


Figure 10: Interaction of Visual Ability and Sound Environment With Respect to Subjective Preference

This graph shows the differences in sound environment preference between subjects from the two visual groups. Visually impaired subjects preferred the tonal environment whereas the sighted users preferred the more musically-based environments.

DISCUSSION

These results have shown that visually impaired users can use the sound environments to perceive depth and position in a 3D application. Although the time for them to locate a position is significantly longer than a sighted user with a display, their accuracy was often very good. This is an important result; it means that a blind user using a tonal sound can perform 3D tasks with the accuracy of a sighted user. This opens a world of possibilities to the visually impaired user.

Comparing both visual ability groups with the sound-only environments showed that visually impaired subjects have a more developed sense of hearing. Sighted users rely on their sight, whereas visually impaired users rely on their sense of hearing. It isn't surprising therefore to find that they out-performed sighted users in this area.

Subjectively, sighted users preferred musical environments whereas visually impaired users preferred the tonal environment. It is thought that this result is due to the way both groups use the sound. Sighted users rely on their vision the most to complete the task. The sound provided is simply an additional cue which sighted users rely on only for depth information. Turning this cue into a musical piece makes it more pleasant for sighted users. Visually impaired users, however, rely only on the sound. The sound that is therefore the most accurate and least distracting is the sound cue of choice. The tonal sound environment was simple and

provided visually impaired users enough information to be very accurate.

When designing these experiments, we expected the orchestral environment to perform better than it did, because it is tailorable (a user could play their favorite piece), and because it addresses a shortcoming of audio environments: the fact that positions are always relative, unless the user has perfect memory for pitch, volume, and balance. We attribute its poor showing to two factors: limited resolution and unfamiliarity. We discuss these two points next.

The orchestral environment had severely limited resolution. Owing to limitations in the audio hardware we could only construct an orchestra consisting of 8 distinct sections arranged in a 4x2 grid. This means that precise positioning had to be determined by the relative volumes of adjacent orchestral sections—a subtle distinction. This lack of resolution certainly contributed to the orchestral environment's poor target error performance.

In addition, the orchestral environment—the arrangement of a set of instruments on a 2D plane—was less familiar to our users than the more conceptually simple tonal or musical environments. In order to determine the effects of familiarity on performance and preference in the three sound environments, we performed a longitudinal study on normally sighted users [8]. This study tested the three environments over three days—a Monday, Wednesday, and Friday of a single week. The longitudinal study showed that performance on all environments (including the orchestral environment) improved over the three days, but the orchestral environment remained the worst performer. However, by the end of the third day it was preferred over the other two environments.

Persistent audio feedback, even when it provides valuable information, is often annoying to users. An environment which minimizes this annoyance is an environment which will actually be used. This appears to be the most important contribution of the orchestral environment for the moment.

In conclusion, sound can help both sighted and visual impaired users in 3D applications. The best sound environment depends on the user's visual ability: sighted users prefer music, visually impaired users prefer tonal. The usefulness of the orchestral environment remains to be shown.

Finally, we have seen that sound-only applications are also possible for both the sighted and visually impaired user, with the visually impaired user being more proficient. This raises the possibility of sighted users operating 3D interfaces in situations where their attention needs to be focussed elsewhere, and opens the world of 3D interfaces to blind users.

ACKNOWLEDGMENTS

Funding for this research was provided by the Natural Sciences and Engineering Research Council of Canada, the University of Waterloo/Institute for Computer Research, and Intel Corporation. We would also like to thank Kirk Reiser of the University of Western Ontario's Blind Lab facility for helping with the experiments.

REFERENCES

1. Di Giano, C.J. & Baecker R.M. Program Auralization: Sound Enhancements to the Programming Environment, in *Proc. Graphics Interface '92* (Vancouver BC, May 1992), CIPS, 44-52.
2. Edwards A.N. The Design of Auditory Interfaces for Visually Disabled Users, in *Proc. CHI'88*, (Washington D.C., May 1988), ACM Press, 83-88.
3. Fitch, W.T. & Kramer, G. Sonifying the Body Electric: Superiority of an Auditory over a Visual Display in a Complex, Multivariate System, in G. Kramer (ed.) *Auditory Display*, Addison-Wesley, Reading MA, 1994, 307-326.
4. Gaver, W.W, Smith, R.B, & O'Shea, T. Effective Sounds in Complex Systems: The Arkola Simulation. in *Proc. CHI'91*, (New Orleans LA, May 1991), ACM Press, 85-90.
5. Hayward, C. Listening to the Earth Sing, in G. Kramer (ed.) *Auditory Display*, Addison-Wesley, Reading MA, 1994, 369-416.
6. Jackson, J.A. & Francioni J.M. Synchronization of Visual and Aural Parallel Program Performance Data, in G. Kramer (ed.) *Auditory Display*, Addison-Wesley, Reading MA, 1994, 291-306.
7. Mansur, D.L, Blattner M.M, Joy, K.I. Sound-Graphs: A Numerical Data Analysis Method for The Blind, in *Proc. 18th Annual Hawaii International Conference on System Sciences*, (Honolulu HI, January 1985), 198-203.
8. Mereu S.W. Improving Depth Perception in 3D Interfaces with Sound, Technical Report No. CS95-35, University of Waterloo, 1995.