Yugoslav Journal of Operations Research
14 (2004), Number 1, 99-115

EXPERIMENTAL EVALUATION OF USER PERFORMANCE IN A PURSUIT TRACKING TASK WITH MULTIMODAL FEEDBACK

Željko OBRENOVIĆ

Faculty of Organizational Sciences, University of Belgrade
Belgrade, Serbia and Montenegro
obren@fon.bg.ac.yu

Received: October 2003 / Accepted: January 2004

Abstract: In this paper we describe the results of experimental evaluation of user performance in a pursuit-tracking task with multimodal feedback. Our experimental results indicate that audio can significantly improve the accuracy of pursuit tracking. Experiments with 19 participants have shown that addition of acoustic modalities reduces the error during pursuit tracking for up to 19%. Moreover, experiments indicated the existence of perceptual boundaries of multimodal HCI for different scene complexity and target speeds. We have also shown that the most appealing paradigms are not the most effective ones, which necessitates a careful quantitative analysis of proposed multimodal HCI paradigms.

Keywords: Multimodal user interfaces, sonification, experimental evaluation, pursuit tracking, human factors**.**

1. INTRODUCTION

Extending user interfaces by using the audio channel is nowadays becoming commonplace. Many applications use different sound effects to extend the perceptual bandwidth of human-computer interaction. One of the most important advantages of sonification is that a computer generated sonic scene leaves the visual field unimpaired, unobstructed, and ready for investigation of the environment for surprises [14, 28]. Sonification can significantly improve the quality of human computer interaction particularly in case of virtual/augmented reality systems. This is very important in a range of the mission critical applications, such as surgical navigation, aircraft navigation and safety, spacecraft docking, night vision mission navigation, and flight navigation and orientation [20, 25]. However, the use of sound modalities is not a universal solution, and wider acceptance of sonification and audio applications will depend on many factors,

such as the quality of user interfaces and how quick a user can effectively learn to use the environment. Exploration into perceptual features of a plethora of sonification paradigms requires lots of research and flexible environments that could be easily customized to suit user or applications needs [22].

In this paper we describe the results of experimental evaluation of user performance in a pursuit-tracking task with multimodal feedback. In a pursuit-tracking test, a moving target moves over the screen under computer control [6]. The operator uses a control device, such as a mouse, to track the target's motion. We enriched feedback about operator's performance, usually given by a tracking symbol in the form of a cursor, with different sound modalities, and explored how the inclusion of these simple audio modes in feedback can improve the accuracy of user's movements. In order to quantify the results of experimental evaluation of selected sonification techniques, we developed a multimodal simulation and training system. In the paper we introduced the quantitative results of three pursuit tracking applications using a combination of acoustic and visual guidance and different background conditions. Experiments have shown that acoustic presentation improved the quality of human-computer interaction and reduced the error during pursuit tracking tasks. Moreover, experiments have shown that benefits do not exist in all conditions, indicating the existence of perceptual boundaries of multimodal HCI for different scene complexity and target speeds.

This paper is organized as follows. Typical issues and survey of previous work of interest to pursuit tracking tasks are given in section 2. Design and implementation of the environment are presented in section 3. Organization of experiments is described in section 4, while analysis of results of experiments is given in section 5. Results and discussion of user performance is given in section 6. Section 7 concludes the paper.

2. PREVIOUS RESEARCH AND BACKGROUND

In this section we describe previous work in two HCI domains of interest for multimodal pursuit tracking tasks:

- *Aimed movement with visual feedback*, where the user receives visual feedback about the position of the target and its performance, and
- *Multimodal presentation*, where the user receives feedback over various modalities, such as vision and sonification.

2.1. Aimed Movements with Visual Feedback

Many of the existing solutions, such as those in pointing and trajectory based human-computer interface (HCI) tasks, have primarily explored visual feedback, and sometimes in rather limited conditions. Visual feedback is very helpful cue in any human motor activity. Relation between aimed movements and vision is fundamental and widely explored topic. There are many empirical laws that model various aspects of human performance. For example, Fitts' law, a psychological model of human movements, explores the visual and haptic feedback in aimed hand movement tasks [17, 1]. Fitts found logarithmic dependency between task difficulty and the time required to complete the movement task. Fitts' law has proven one of the most robust, highly cited, and widely adopted models, and has been applied in diverse settings – from under a microscope to

under water activities. Initially demonstrated on one-dimensional tasks Fitts' law was extended to more complex two-dimensional and dynamic tasks, as well as to high precision tasks [18, 11]. Based on Fitts' law, the International Standards Organization (ISO) has developed the Part 9 (Non-keyboard Input Device Requirements) of the ISO standard 9241 – Ergonomic Requirements for Office Work with Visual Display Terminals [12].

2.2. Multimodal presentation and sonification

Usage of visualization as a feedback for movements, although effective, is not always practically feasible or even possible. When the visual field is overwhelmed, audio feedback may be extremely useful [10]. Sonification is the second most important presentation modality after visualization [4, 16]. In human-computer interaction audio presentation can be effectively used for many purposes. Voice and sound guide the attention and give additional value to the content through intonation [26]. The sound dimension offers an alternative to reading text from the computer screen, what can be especially important for visually impaired users.

Relationship between visualization and sonification is itself a complex design problem, due to the nature of the cognitive information processing. Efficiency of sonification, as acoustic presentation modality, depends on other presentation modalities [3]. Some characteristics of visual and acoustic perception, such as spatio-temporal resolution, are complementary. Therefore, sonification naturally extends visualization. Well-designed multimodal systems integrate complementary modalities to yield a highly synergistic mix in which the strengths of each mode are used to overcome weaknesses in the other. Such systems can function more robustly than unimodal systems [5, 23].

Human vision and audition perception profit from different aspects of the various stimuli coming from the outside environment. We can hear in all directions, but we must turn to see behind us, so it is usually said that "the ears guide the eyes" [8]. However, visual and aural perceptions have their strengths and weaknesses, which have to be taken into account when designing user interfaces. For example, while it is faster to speak than to write, it is faster to read than to listen to speech [7]. Table 1 presents some guidelines about the use of audio and visual display, according to the type of the message being presented [9, 7].

Table 1: Guidelines about the use of audio and visual displays [9]

	Use auditory presentation if:	**Use visual presentation if:**
1.	*The message is simple.*	*The message is complex.*
2.	*The message is short.*	*The message is long.*
3.	*The message will not be referred to later.*	*The message will be referred to later.*
4.	*The message deals with events in time.*	*The message deals with location in space.*
5.	*The message calls for immediate action.*	*The message does not call for immediate action.*
6.	*The visual system of the person is overburdened.*	*The auditory system of the person is overburdened.*
7.	*The receiving location is too bright or dark — adaptation integrity is necessary.*	*The receiving location is too noisy.*
8.	*The person's job requires him to move about continually.*	*The person's job allows him to remain in one position.*

Humans are naturally skilled to perceive visual and sound clues simultaneously. For example, listening to speech is often complemented with naturally evolved lip-reading skills, which allows more efficient understanding of conversation in a noisy environment. Experiments also suggest that even simple speechless audio cues can further improve unimodal visual environments [22].

Many systems have tried to exploit the advantages of sound presentation. For example, tactical audio systems are a special class of multimodal interfaces which use audio feedback to facilitate the precise and accurate positioning of an object with respect to some other [15]. Tactical audio has valuable applications in the field of surgery. The use of tactical audio feedback enables the surgeon to effect a precise placement by enhancing his/her comprehension of the three-dimensional position of a surgical instrument with respect to some predetermined desired position within the patient's body [28]. Sonification could be used to facilitate insight into complex phenomena. We have applied sonification of brain electrical activity to improve the presentation of complex spatio-temporal patterns of brain electrical activity [15].

3. THE DESIGN OF A MULTIMODAL SIMULATION SYSTEM

Although there have been many notes on usefulness of multimodal interaction, many issues on usage of sound and other modalities, as well as on their integration still need to be solved. Therefore, we have developed a multimodal simulation and training environment for experimental evaluation of different sonification paradigms. The environment allows recording of relevant parameters of user interaction during pursuit tracking tasks, and supports quantitative evaluation of effectiveness of sonification methods and assess user's learning curve in a large population of users [22]. In addition, our goal was to investigate suitability of standard PC software and hardware modules for efficient implementation of tactical audio applications.

3.1. The high-level model of the Multimodal Simulation System

User interfaces can be viewed as one-shot, higher-order messages sent from designers to users [24]. While designing a user interface the designer defines an interactive language that determines which messages will be included in the interaction. Hence, we firstly present the structure of these high-order messages in the form of a high-level model of our environment. The main concept of our model is the concept of a computing mode. We defined a computing mode as a form of interaction that was designed to engage some human capabilities, with designer's aim of sending some message to a user. We have classified messages that a mode can send in three main categories: sensual, perceptual, and cognitive. A sensual message is a low-level stimulus effect aimed to excite some parts of human sensory apparatus. A perceptual message engages some of human perceptual skills such as pattern recognition or perception of three-dimensional cues. A cognitive message engages more complex human capabilities such as memory, attention or cognitive chunking. A complex computing mode integrates other modes to create simultaneous use of various modalities, while a simple mode represents a primitive form of interaction. We have defined input and output types of simple computing mode. Input computing mode requires some user devices to transfer

human output into a form suitable for computer processing. Output computing mode presents data to the user. A presentation can be static or dynamic.

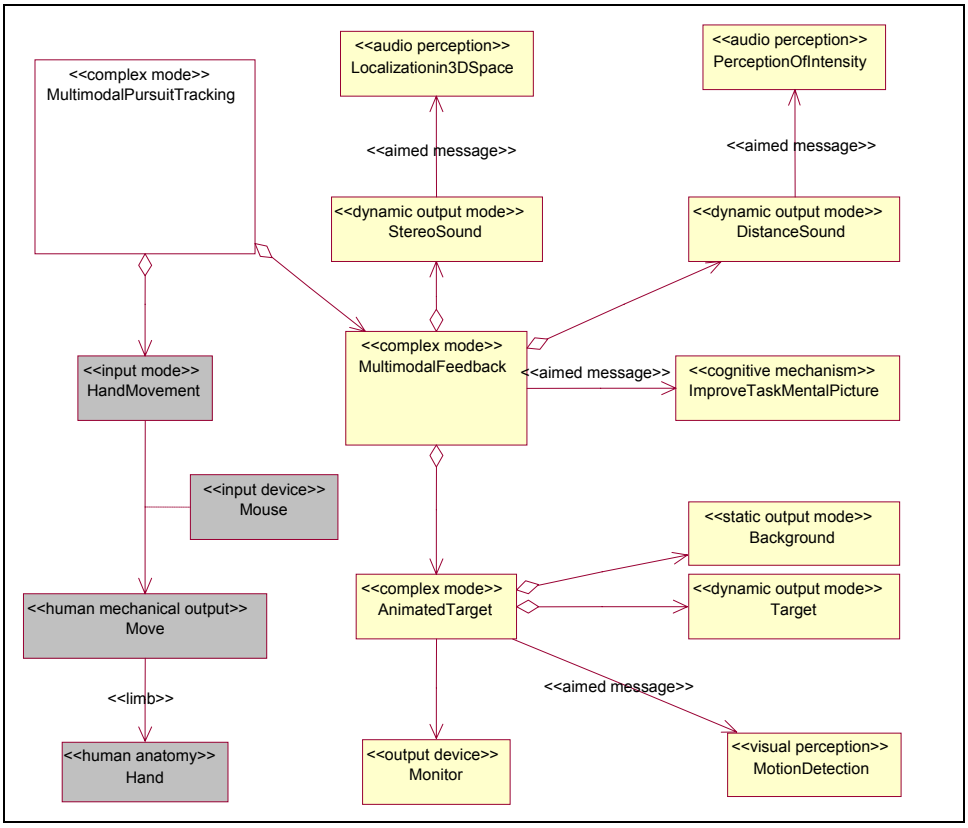

Figure 1: A model of multimodal pursuit tracking environment described with UML

Figure 1 shows the high-level model of our multimodal pursuit-tracking environment. As the figure shows, pursuit tracking is a complex mode that integrates hand movement human output, and multimodal feedback. We used mouse as an input device. Multimodal feedback is a complex output mode that integrates a visual presentation, and two audio modes. A visual representation integrates static background presentation and animated target in order to attack user's visual motion detection perceptual mechanism. The first audio mode is designed to produce three-dimensional stereo effect using stereo cues [21, 19], while the second audio mode warns a user about the tracking error by changing the intensity of sound.

3.2. The Architecture

The architecture of the multimodal test environment is shown in Figure 2. The environment consists of an interaction space and a control interface. Main interaction between the user and the environment occurs in the interaction space, where the user

tracks an animated target on a 2D polygon. The environment is designed flexibly in order to allow easy plug-and-play addition of new graphical and acoustic modes.

The interaction space consists of a multimodal integration module and path interpolator. Multimodal integration module does the multimodal presentation of data. We used one graphics mode in the form of animated circle object. In experiments a user simply tries to "cover" the center of the target using cursor. We also used two different sound modes described in the previous section. The user could use any combination of these three modes. Path interpolator calculates new target position on each timer tick using discrete set of coordinates to provide "smooth" trajectory in space/time.

In the control interface the user can set various parameters including multimodal combination, speed, and file parameters. It also maintains generation of a path from a text file, and writes the results to the file. A path file is a simple text file that contains trajectory of an object as a discrete set of *X* and *Y* coordinates. The result file is an XML file that contains samples of user cursor coordinates together with the samples of object coordinates. The result file also contains a description of experimental conditions.

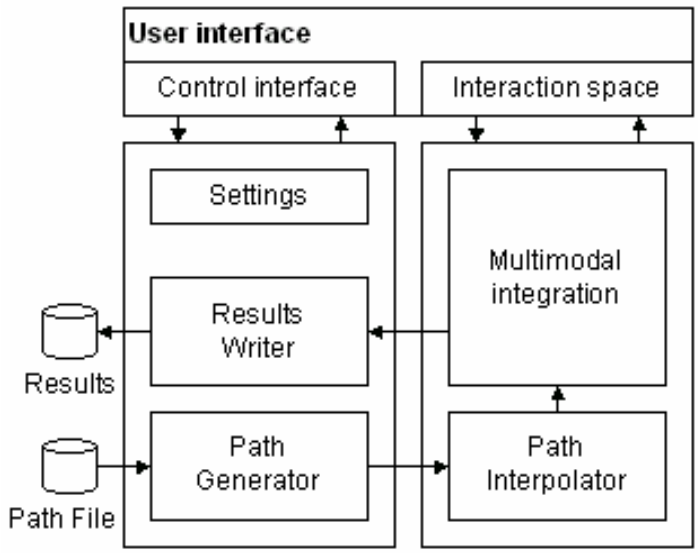

Figure 2: The architecture of our environment.

The environment is based on commercially available technology, implemented using Java3D package [27] and standard Java components. This approach allows transparent execution of the environment on various platforms including stand-alone workstations, as well as distributed Web environments in the form of Java applet. The Java3D package is primarily used for acoustic 3D effects.

The environment can be used in two modes: *training* and *experiment* mode. In the training mode a user can arbitrarily set multimodal combination, speed of the object, start and duration of a session. This mode is particularly suitable for training of users before the experiment, and for pilot testing of new modes. *Experiment* mode is

designated for predefined tests, where users are not allowed to change any of the parameters in the environment. All parameters are preset in a file. Each experiment is defined as a sequence of tests. For each test we define speed of the target object, duration of interaction, length of pause between two tests, as well as presentation modes used in the test. All of these parameters are stored in result file, and later used in analysis.

3.3. The Multimodal Presentation Environment

We have developed three variants of the environments in order to facilitate testing of various interaction complexities:

- **E1 - Homogenous background**
  The simplest environment presents the circular object that moves over homogenous single-color background (Figure 3).

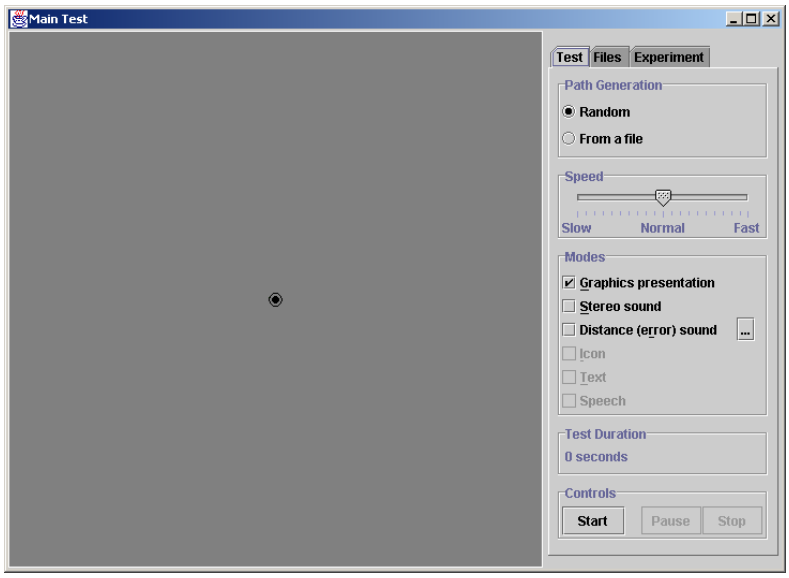

Figure 3: The user interface of the environment with static background.

- **E2 - Static heterogeneous background**
  In this environment the object moves over static heterogeneous background, which effectively adds visual noise to graphical presentation. Static heterogeneous background was simulated with a static picture of the map (Figure 4).

- **E3 - Moving background**
  In this variant of the environment the object is a part of static heterogeneous background. The object does not move relative to the background. Instead, the whole background, including the object, moves across the screen. For example, this way of interaction can be illustrated with aircraft simulation where you have to pursuit a static object on the ground while driving an aircraft.

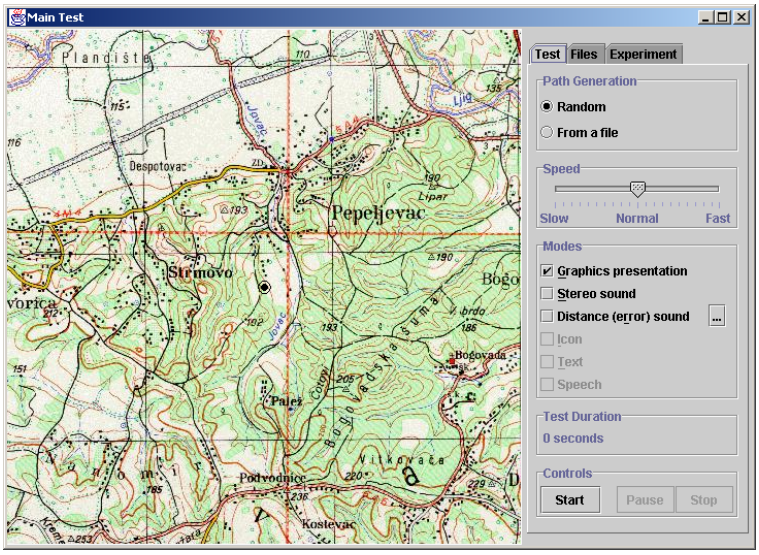

Figure 4: The user interface of the environment with static heterogeneous background

4. EXPERIMENTS

Target tracking experiments were conceived to assess user performance in two-dimensional pursuit tracking applications. We organized three experiments, each with one variant of the environment, as described in previous section. Experimental setup, parameters, methods and procedures were the same for all of the tests.

4.1. Method

The task of a user was to pursuit a moving object on the screen as close as possible. Target object was positioned according to a predefined trajectory taken from a file, and coordinates of both the target object and the cursor were written to the result file for off-line processing.

We created four multimodal combinations (MMCs) (see Table 2). Each of the combinations was tested with four object speeds (see Table 3).

Table 2: Multimodal combinations (MMCs) used in the experiments

	Graphics mode	Distance sound	Positional sound
MMC1	+	–	–
MMC2	+	+	–
MMC3	+	–	+
MMC4	+	+	+

Table 3: Average speeds of the target object in the experiments

Speed ID	Average speed (pixels per second)
1	50
2	100
3	200
4	300

Experiments ran on an 800 MHz Intel Pentium III desktop PC with 128 MB of RAM with Windows 2000 Professional operating system. Computer had S3 Incorporated Trio3D/2X graphics card, Samsung 15" Sync Master 550s monitor and integrated CMI 8738/C3DX PCI audio device. Participants used headphones for better spatial sound perception. Screen resolution was set to 800x600 pixels. Dimensions of the interaction space were 500x500 pixels. Mouse motion sensitivity was set to medium. Frame rate for the first two experiments was about 50 frames per second, and for experiment E3 it was about 30 frames per second. The difference in frame rate among the experiments is a consequence of different graphical complexity of the scenes, as the moving background requires more time to render.

4.2. Participants

Nineteen persons participated in experiments. Participants were unpaid male volunteers from military academy. Age of the participants ranged from 19 to 30 years. All participants had normal or corrected-to-normal vision and normal hearing, and were right-handed. All of them have regularly used GUIs and a mouse.

4.3. Experimental Procedure

All three experiments followed the same procedure. Participants were first asked to fill a questionnaire that assessed their prior experience with computers and mouse use. Then, participants practiced interaction with the environment in the training mode. During this period all the speeds, as well as multimodal combinations covered in experiment were presented to the user.

After warming up, participants started the experiments. In order to eliminate effects of learning on results, a sequence of experiments for participants in the group was not the same. Each experiment consisted of 16 tests (4 multimodal combination x 4 speeds). Each test lasted for 35 seconds. Total duration of one experiment for each participant was about 10 minutes. This relatively short duration of each test is chosen since longer sessions had resulted in strong finger and wrist fatigue of participants. The experiment started with lowest speed, where all combinations were tested. After that the speed was increased. Before new test at higher speed, participants were able to make a pause. We encouraged participants to rest before proceeding to the next level.

After experiment completion, participants were asked to rank multimodal combinations from one to four, where one represents the multimodal combination they found the most helpful during interaction.

5. ANALYSIS

Positions of the target object and the cursor were collected directly by the software. The data were then prepared for statistical analysis by computing values of a tracking error for each trial, and average tracking error for each speed and multimodal combination.

We excluded trials of participants who reported troubles during the sessions such as lack of attention due to some distraction or problems with handling a mouse. We excluded approximately 5% of all trials.

For each sample in all experiments we calculated error as Euclidean distance between user cursor and the target object. After that, we calculated mean error and standard deviation for each multimodal combination and speed. We also calculated benefit ξ of using multi modal HCI compared to pure graphics mode (MMC1):

$$\xi = 1 - \frac{\text{error}_{\text{AUDIO}}}{\text{error}_{\text{MMC1}}}$$

where $\text{error}_{\text{MMC1}}$ represents an average error for first multimodal combination (MMC1), and $\text{error}_{\text{AUDIO}}$ is an average error for multimodal combinations with multi modal combinations (MMC2–MMC4) running *at the same speed*. We defined three relative errors, each for one multimodal combination with sound nodes. We then calculated paired T-test between each of the multimodal combinations with sound modes (MMC2–MMC4) and pure graphics mode (MMC1) for the same target speed.

6. RESULTS AND DISCUSSION

6.1. Results

Results of the analysis for all experiments are summarized in Table 4. Results are grouped by speed and multimodal combination. Average errors in all three experiments are shown in the first three columns, while benefits of using audio modes compared to pure graphical mode for the same speed are shown in the last three columns.

Paired t-tests between multimodal combinations with audio modes (MMC2 – MMC4) and pure graphics mode (MMC1) for the same speed revealed significant differences in average error across ten tests. In the first experiment only significant effect existed for the second target speed, for MMC3 $(p = 0.009)$ and MMC4 $(p = 0.002)$. Both combinations use spatial sound mode.

In the second experiment significant change can be seen at the second and third target speed. For the second speed significant effect existed for multimodal combination with stereo position sound mode MMC3 $(p \le 0.002)$ and MMC4 $(p \le 0.0009)$. At the third target speed all multimodal combinations had significantly different average errors, MMC2 $(p \le 0.001)$, MMC3 $(p \le 0.01)$, and MMC4 $(p \le 0.002)$.

In the third experiment significant effect existed for the first and second speed. For the first speed MMC2 had significantly different average values $(p \le 0.005)$. For the

second speed multimodal combination with stereo distance sound mode had significantly different average errors MMC3 ($p \leq 0.0004$) and MMC4 ($p \leq 0.002$).

Table 4: Average distance errors with standard deviations and benefits multi-modal combinations for all three experiments (E1, E2, E3). Values marked with '*' are statistically significant.

		Average error in pixels (SD)			**Benefit in %**		
Speed	**MMC**	**E1**	**E2**	**E3**	**E1**	**E2**	**E3**
1	1	7.40 (2.42)	7.18 (1.60)	7.69 (1.45)			
	2	7.03 (2.05)	7.45 (2.40)	**6.55 (1.29)***	5.3	-3.7	17.5
	3	7.47 (2.04)	7.51 (1.68)	7.92 (2.14)	-0.9	-4.4	-2.8
	4	7.27 (3.36)	7.57 (1.92)	7.69 (1.58)	1.7	-5.2	0.1
2	1	12.81 (3.30)	13.74 (3.39)	12.37 (1.68)			
	2	12.14 (2.87)	12.99 (3.91)	11.79 (1.48)	5.6	5.8	4.9
	3	**11.17 (2.22)***	**11.68 (2.42)***	**10.82 (1.73)***	14.7	17.6	14.3
	4	**10.76 (1.75)***	**11.96 (2.74)***	**11.44 (2.00)***	19.1	14.8	8.2
3	1	21.21 (2.60)	24.20 (2.96)	23.19 (2.84)			
	2	21.38 (3.33)	**21.57 (3.33)***	22.30 (3.71)	-0.8	12.2	4.0
	3	20.25 (3.49)	**22.45 (4.04)***	22.31 (3.42)	4.7	7.8	3.9
	4	20.24 (3.50)	**22.09 (4.10)***	22.84 (4.63)	4.8	9.5	1.5
4	1	33.45 (3.79)	36.75 (4.21)	40.18 (6.79)			
	2	34.20 (5.55)	37.24 (5.48)	38.96 (4.86)	-2.2	-1.3	3.1
	3	33.72 (3.79)	35.61 (3.86)	38.92 (6.27)	-0.8	3.2	3.2
	4	33.91 (4.78)	36.70 (5.30)	39.92 (8.86)	-1.4	0.1	0.6

Figures 5, 6, and 7 illustrate dependencies among speed, multimodal combination and benefits of multimodal HCI for all three experiments. Values marked with '*' are statistically significant.

It could be seen that the spatial audio mode significantly improved pursuit precision for the second target speed, since benefits of MMC3 and MMC4 combinations were significant and positive for all three experiments. Benefits of MMC3 and MMC4 were in the first experiment 15% and 19%, in the second experiment 18% and 15%, and 14% and 8% in the third experiment. The second experiment exhibited benefits for all audio modes for the third speed although these benefits were smaller than in case of the second target speed (12%, 8%, and 10%). The third experiment has shown significant benefit for distance mode (MMC2) in the first speed (18%). Benefits were smaller and not statistically significant for other tests.

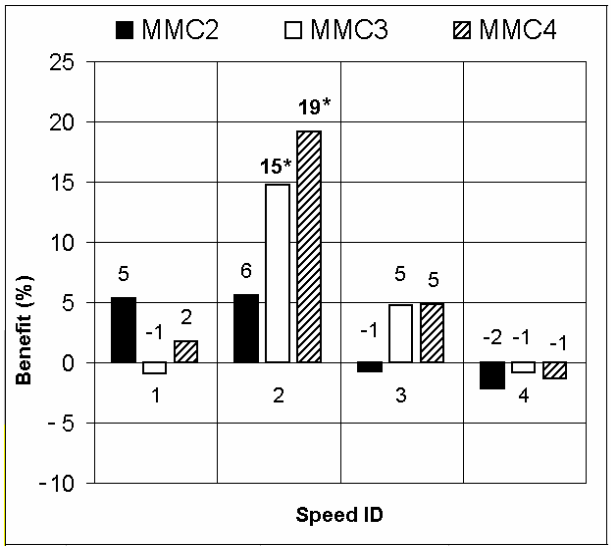

Figure 5: Benefits of multimodal HCI in the first experiment (Homogenous background)

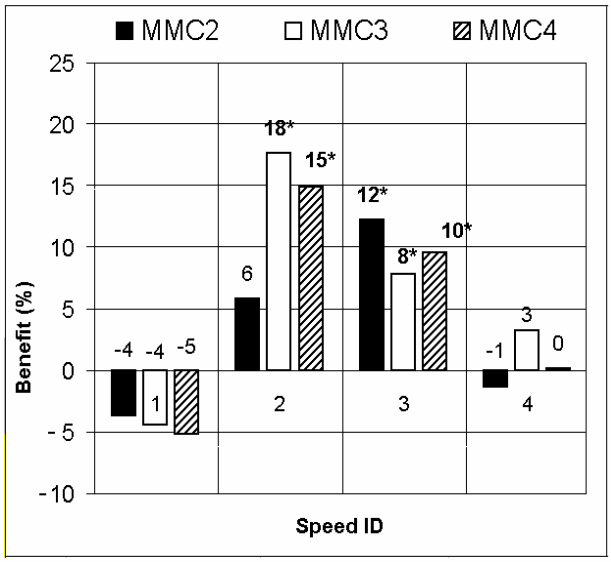

Figure 6: Benefits of multi-modal HCI in the second experiment (Static heterogeneous background)

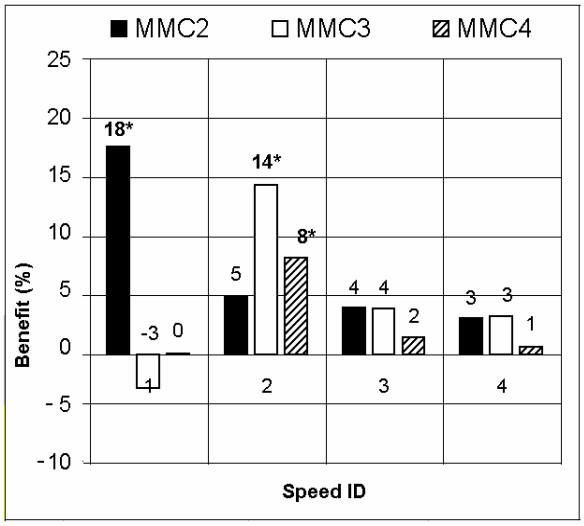

Figure 7: Benefits of multi-modal HCI in the third experiment (Moving background)

Results of user's subjective ranking of multimodal combinations after the experiment are presented in Table 5.

Table 5: User's evaluation of multimodal combinations. Smaller grade represents better rank

Rank	Multimodal combination	Average mark
1	**MMC2**	1.55
2	**MMC4**	2.55
3	**MMC1**	2.73
4	**MMC3**	3.00

6.2. Discussion

Experiments have shown that audio modalities could significantly improve the quality of human-computer interaction. Users were more precise when they used multimodal combinations with acoustic modes, but benefits do not exist in all conditions. Experiments have shown that benefits exist only for one or two speeds, indicating perceptual boundaries for efficient use of audio modalities.

There are many perceptual and physiological factors that have to be taken into consideration when discussing the results. As the vision is the main sensory channel for most of the humans, it is important to note some of the characteristics of human visual apparatus. For example, human eye is not uniformly sensitive to details and motion. Fovea and parafoeva eye sensor regions are very sensitive to details in the central part of

the visual field. As a consequence of this disproportion, no less than 80% of the visual cortex is involved in processing of less than 10° of visual field. In addition, humans' visual system has specialized neural circuitry for motion perception. These low-level pre-attentive vision processes enable quick motion perception [13]. Motion is very effective at making one takes notice and is one of the strongest visual appeals to attention. As a result, in a visually crowded environment the moving objects stand out clearly from the rest [2].

In our experiments, as the dimension of the target was small, we expected the users to have been visually tracking the animated target with fovea area, which means that they were able to visually detect even small changes near the target. However, having in mind the dimensions of the presentation screen, the distance of the user from the screen, and speed of the target, it is possible that the distance between the target and the cursor would include parafovea and peripheral vision for very high speeds, too. In addition, animated graphical presentation of the target in our environment activates visual motion-detection perceptual mechanism.

Having in mind the above mentioned properties of visual perception, it can be seen that great part of users' visual apparatus was involved in processing of details in central part of users' visual field. However, this means that the rest of the interaction space was weakly covered with vision, complicating maintenance of the mental picture of the overall context. We assume that this explains statistically significant benefits of stereo sound: *the vision gave us a good sensitivity for details, while stereo sound provides us with a context of the overall interaction space*. This, once more, illustrates advantages of using audio modes, because a computer generated sonic scene leaves the visual field unimpaired, unobstructed, and ready for investigation of the environment for surprises. It is also worth noting that multimodal combination with stereo positioning sound (MMC3) had the most stable behavior in all experiments. Figure 8 illustrates dependency between speed and benefits of MMC3 for all experiments. This result indicates that stereo sound can bring consistent benefit in various conditions, what makes it more appropriate for implementation in pursuit tracking applications.

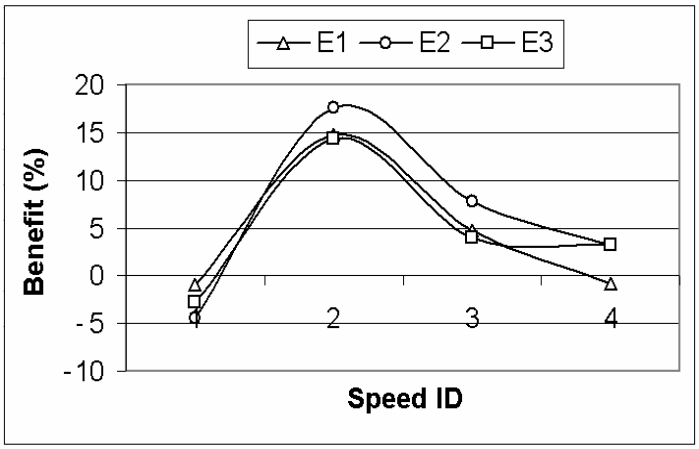

Figure 8: Benefits of spatial stereo sound for all experiments and different target speed

Although main objective of our research was to investigate the benefits of the use of audio modes, we made some analysis of average errors for pure graphics mode. Normally, higher speed produces higher error (Figure 9). It could be seen that linear trend-line quite well approximates the average error ($R^2 > 0.99$). In other words, this means that in our environment average error for pure graphical feedback can be roughly estimated by dividing the distance that the target traverses in one second by constant 8. However, due to small number of examined speeds, these claims will be the topic of a future study.

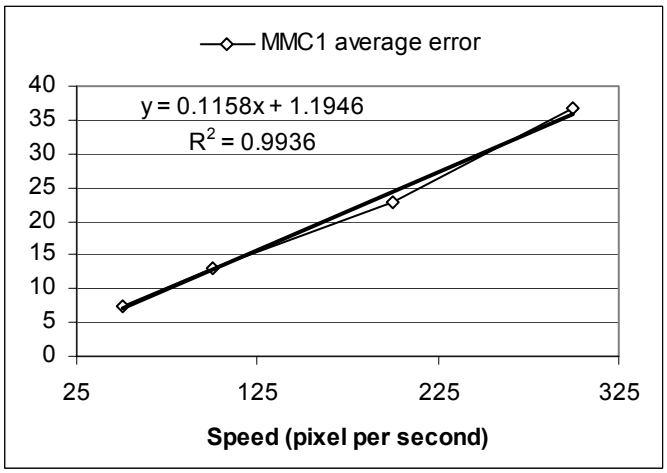

Figure 9: Dependency between speed and average error. Bold line represents the trend-line

Results have shown that the complexity of visual presentation, simulated with different backgrounds, also influences benefits of audio modalities, but not in a statistically significant way. This suggests extension of experiments to also include other visualization paradigms, in order to find the relationship between parameters of visual presentation and benefits of audio modalities.

It is interesting to note that there exists a gap between user's grades and experimental results. Users ranked the highest audio distance mode (MMC2), which alone showed statistically significant benefits only twice. On the other hand, users ranked the worst the stereo position sound (MMC3) which alone showed statistically significant benefits four times. The analysis of users' grades indicates that users are not always aware of benefits of spatial audio mode.

Our goal was to investigate how much the standard PC software and hardware modules are suitable for efficient implementation of tactical audio. For that reason, our experiments ran on relatively inexpensive equipment, using the Java3D packet in spite of some limitations for sounds effects.

7. CONCLUSION

In this paper we have described the results of experimental evaluation of user performance in a pursuit-tracking task with multimodal feedback. Usage of audio modes as an aid to achieve precise positional control in pursuit tracking applications, exhibits a significant potential for improved user performance and better quality of human-computer interaction. In order to evaluate various interaction paradigms we developed a multimodal simulation system for testing of different sonification and visualization paradigms. Our experimental results indicate that audio can significantly improve the accuracy of pursuit tracking. While the vision gives us a good sensitivity for details in local space, three-dimensional sound can provide a context of the overall interaction space. In addition, the experiments have shown that benefits are limited by user perception, indicating the existence of perceptual boundaries of multimodal HCI for different scene complexity and target speeds. Also, the experiments in our environment have shown that acoustic modes improve human computer interaction even with relatively poor audio quality of the PC system that we used.

Experiments have also shown that users are not aware of quantitative benefits of applied audio modalities. We have shown that the most appealing paradigms are not always the most effective ones, which necessitates a careful quantitative analysis of proposed multi-modal HCI paradigms. Therefore, it is necessary to experimentally evaluate audio user interfaces, as we cannot rely on users' subjective estimations.

Further work will include extension of experiments to include other visualization paradigms, in order to find relationship between parameters of visual presentation and benefits of audio modalities. Future research will also include a larger number of experiments, analysis of users learning curves, and multimodal guidance of a truly portable, PDA based, guidance system.

Acknowledgments: The authors would like to acknowledge Dragan Gasević and Miroslav Havram for their valuable help in organization of the experiments.

REFERENCES

[1] Akamatsu, M., MacKenzie, I.S., and Hasbrouq, T., "A comparation of tactile, auditory, and visual feedback in a pointing task using a mouse-type device", *Ergonimics*, 38 (1995) 816-827.
[2] Arnheim, R., *Art and Visual Perception*, The University of California Press, Berkeley, 1974.
[3] Barnard, P.J., and May, J. (eds.), *Computers, Communication and Usability: Design Issues, Research and Methods for Integrated Services*, North Holland Series in Telecommunication, Amsterdam, Elsevier, 1993.
[4] Begault, D.R., *3D Sound for Virtual Reality and Multimedia*, Academic Press, Boston, 1994.
[5] Blattner, M.M., and Glinter, E.P., "Multimodal integration", *IEEE Multimedia, IEEE CS Press*, 3 (4) (1996) 14-24.
[6] Buxton, W., "Touch, gesture & marking", in: R.M. Baecker, J. Grudin, W. Buxton, S. Greenberg (eds.), *Readings in Human Computer Interaction: Toward the Year 2000*, Morgan Kaufmann Publishers, San Francsco, 1995.
[7] Buxton, W., "Speech, language & audition", in: R.M. Baecker, J. Grudin, W. Buxton, S. Greenberg (eds.), *Readings in Human Computer Interaction: Toward the Year 2000*, San Francisco, Morgan Kaufmann Publishers, 1995.

[8] Cook, P.R., "Virtual worlds, real sounds", *IEEE Computer Graphics and Applications, IEEE CS Press*, 22 (4) (2002) 22.
[9] Deatherage, B.H., "Auditory and other sensory forms of information presentation", in: H.P. Van Cott, R.G. Kinkade (eds.), *Human Engineering Guide to Equipment Design (Revised Edition)*, U.S. Government Printing Office, Washington, 1972.
[10] DiFilippo, D., and Pai, D.K., "The AHI: an audio and haptic interface for contact interactions", *Symposium on User Interface Software and Technology*, San Diego, California, United States, ACM Press, New York, NY, USA, 2000, 149-159.
[11] Guiard, Y., Beaudouin-Lafon, M., and Mottet, D., "Navigation as multiscale pointing: extending Fitts model to very high precision tasks", *Conference on Human Factors and Computing Systems*, Pittsburgh, Pennsylvania, United States, ACM Press, New York, 1999, 450-457.
[12] ISO ISO/TC 159/SC4/WG3 N147: Ergonomic requirements for office work with visual display terminals (VDTs) - Part 9 - Requirements for non-keyboard input devices, International Organisation for Standardisation, May 25, 1998.
[13] Ivry, R.B., and Cohen, A., "Dissociation of short- and long-range apparent motion in visual search", *Journal of Experimental Psychology: Human Perception and Performance*, 16 (2) (1990) 317-331.
[14] Jovanov, E., Wagner, K., Radivojevic, V., Starcevic, D., Quinn, M., and Karron, D., "Tactical audio and acoustic rendering in biomedical applications", *IEEE Transactions on Information Technology in Biomedicine*, 3 (2) (1999) 109-118.
[15] Jovanov, E., Starcevic, D., and Radivojevic, V., "Perceptualization of biomedical data", in: M. Akay, A. Marsh (eds.), *Information Technologies in Medicine, Volume I: Medical Simulation and Education*, John Wiley and Sons, 2001.
[16] Kramer, D. (ed.), *Auditory Display, Sonification, Audification and Auditory Interfaces*, Addison-Wesley, Reading, MA, 1994.
[17] MacKenzie, I.S., "Fitts' law as a research and design tool in human computer interaction", *Human -Computer Interaction*, 7 (1992) 91-139.
[18] MacKenzie, I.S., and Buxton, W., "Extending Fitts law to two-dimensional tasks", *Conference on Human Factors and Computing Systems – CHI 92*, Monterey, California, United States, ACM Press, New York, 1992, 219-226.
[19] Martens, W.L., "Psychophysical calibration for controlling the range of a virtual sound source: Multidimensional complexity in spatial auditory display", *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland, 2001, 231-234.
[20] Murray, J., "Wearable computers in battle: Recent advances in the Land Warrior System", *Proc. of the Fourth International Symposium on Wearable Computers (ISWC'00)*, Atlanta, 2000.
[21] Neuhoff, J.G., "Perceiving acoustic source orientation in three-dimensional space", *Proceedings of the 2001 International Conference on Auditory Display*, Espoo, Finland, 2001, 231-234.
[22] Obrenovic, Z., Starcevic, D., and Jovanov, E., "Experimental evaluation of multimodal human computer interface for tactical audio applications", *Proceedings of IEEE International Conference on Multimedia and Expo - ICME 2002*, Lausanne, Switzerland, 2002, 29-32.
[23] Oviatt, S.L., "Ten myths of multimodal interaction", *Communication of the ACM*, 42 (1999) 74-81.
[24] Prates, R., De Souza, D., and Barbosa, S., "A method for evaluating the communicability of user interfaces", *ACM Interactions*, 7 (1) (2000) 31-38.
[25] Satawa, R., "Future technologies for medical applications", in: M. Akay, A. Marsh (eds.), *Information Technologies in Medicine, Volume I: Medical Simulation and Education*, John Wiley and Sons, 2001.
[26] Schär, S.G., and Krueger, H., "Using new learning technologies with multimedia", *IEEE Multimedia, IEEE CS Press*, 7 (3) (2000) 40-51.
[27] Sowizral, H., Rushforth, K., and Deering, M., *The Java 3D API Specification, Second Edition*, Addison-Wesley Pub Co, 2000.
[28] Wegner, K., and Karron, D., "Surgical navigation using audio feedback", in: K.S. Morgan et al (eds.), *Medicine Meets Virtual Reality: Global Healthcare Grid*, IOS Press, Washington, 1997, 450-458.