Comparison of effectiveness in using 3D-audio and visual aids in identifying objects in a three-dimensional environment

Karl Åbom
This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Bachelor of Science in Digital Game Development. The thesis is equivalent to 10 weeks of full time studies.

Contact Information:
Author(s):
Karl Åbom
E-mail: kaai10@bth.se

University advisor:
Dr. Veronica Sundstedt
Department of Creative Technologies

Faculty of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57
Acknowledgements

The author would like to thank Dr Veronica Sundstedt, for her guidance and advice during the project, Erik Norlen, whose expertise in the Unity engine was an invaluable source of guidance when writing the code for the experiment software, and Erik Wadstein, who helped to find testplayers to carry out the experiment.

Karl Åbom
Abstract

**Context:** Modern commercial computer games use a number of different stimuli to assist players in locating key objects in the presented Virtual Environment (VE). These stimuli range from visual to auditory, and are employed in VEs depending on several factors such as gameplay design and aesthetics.

**Objectives:** This study compares three different localization aids in order to evaluate their effectiveness in VEs.

**Method:** An experiment is carried out in which testplayers are tasked with using audio signals, visual input, as well as a combination of both to correctly identify objects in a virtual scene.

**Results:** Results gained from the experiment show how long testplayers spent on tests which made use of different stimuli. Upon analyzing the data, it was found that that audio stimulus was the slowest localization aid, and that visual stimulus and the combination of visual and auditory stimulus were tied for the fastest localization aid.

**Conclusions:** The study concludes that there is a significant difference in efficiency among different localization aids and VEs of varied visual complexity, under the condition that the testplayer is familiar with each stimuli.

**Keywords:** 3D-games, Auditory localization, Localization aids
| Figure 2.1  | Audio test | 5 |
| Figure 2.2  | Visual test | 6 |
| Figure 2.3  | Audio+Visual test | 7 |
| Figure 2.4  | Complexity level 1 (Displays easiest VE) | 8 |
| Figure 2.5  | Complexity level 2 (Displays normal VE) | 9 |
| Figure 2.6  | Complexity level 3 (Displays hardest VE) | 9 |
| Figure 2.7  | Test environment | 11 |
| Figure 2.8  | Close-up of testers environment | 12 |
| Figure 3.1  | Results table from first round of testing | 15 |
| Figure 3.2  | Results table from second round of testing | 16 |
| Figure 3.3  | Paired T-test values for each test | 17 |
| Figure 3.4  | Average times spent on each test | 18 |
| Figure 3.5  | Average times to click a object per test | 18 |
| Figure 3.6  | Results from post-test questionnaire | 19 |
# Contents

Abstract 1
Figures listing II

## Chapter 1 Introduction 1

1.1 Related work 2
1.2 Objective 2
1.3 Research questions and hypothesis 3

## Chapter 2 Method 4

2.1 Experiment software description 4
2.2 Counterbalancing the test 10
2.3 Preparation and setup 11
2.4 Participants 12
2.5 Procedure 12

## Chapter 3 Results 14

3.1 Test data 14
3.2 Difficulty as complexity of VE increases 18
3.3 Perceived difficulty 19
3.4 Discussion 19

## Chapter 4 Conclusions 21

4.1 Evaluating the results 21
4.2 Future work 22

## Appendix

A.1 User consent document 23
A.2 Pre-questionnaire 25
A.3 Description and instructions for tests 26
A.4 Post-questionnaire 27

## References 28
Chapter 1

Introduction

Chapter content:

1.1 Related work
1.2 Objective
1.3 Research Questions and Hypothesis

Commercial computer games often rely on different stimuli or aids in helping a player identify objects in their VE. As computer games get more complex VEs, different techniques have been introduced to assist players in identifying these objects. (These objects may be dangers around the player, which object in a scene to approach, etc). At times visual aids are used to indicate how a player should proceed (Such as color contrast, lighting [8]), other times audio aids are given and in the case of some computer games, simple haptic feedback in the form of compatible game controllers are all used as ways to relay information to the player. An appropriate aid is often given in the context of the VE; For example, in a horror game the game designers may use audio aids of something approaching the player from behind in order to set the mood for the game.

This study compares the visual and auditory use of localization cues in order to gauge how fast a player can identify a point of interest in a three-dimensional environment created using a premade engine, Unity [1]. To achieve this, a computer game-styled test made in this environment is used. By having testplayers take on progressively more complex tests we gauge users ability to identify objects in a 3D-environment. By changing some of the difficulty parameters in the test, such as the number of potential target objects and the distance between these, a variation in the difficulty of the test will be achieved. Using auditory and visual aids similar to those employed in commercial computer games, the test monitors player performance during these tests. By then comparing how fast and accurate a user is in completing the test using his or hers allotted cues, it is hoped to gain a deeper understanding on how efficient auditory and visual signals are in communicating the position of an object in a 3D-scene to a user.
1.1 Related work

In three-dimensional computer games, there are many different kinds of localization aids. The one that are most relevant to this thesis are 3D-arrows [2] 3D-sound [3]. Two examples of commercial games which employ 3D-arrows as visual aids are Xenoblade chronicles [15] and Crazy taxi [16].

Similar research which study the use of auditory-visual localization performance in VEs exist. For example, some research focus on what attracts players visual attention and monitors this using eye-tracking technology. Specifically, Sundstedt et al. [12] study visual attention in games using eye-tracking. In the study, testplayers were tasked with steering an object through a maze with several distracting objects present while gaze behavior was monitored with the help of eye-tracking technology. On the subject of audio in 3D-games Grimshaw et al. [13] describe a conceptual framework for the design and analysis of audio in first-person shooter games. Their work highlight that the use of sound in a first-person shooter games contribute to player immersion in the game world. In addition, in his presentation on Modern Audio Technologies in Games, Menshinkov [14] looks closer on 3D-sound and its similarities and differences from multichannel solutions. Khoa-Van Nguyen et al. [5] study whether or not auditory-visual VEs is capable of delivering accurate visual and auditory cues in a static situation. Their results confirm the importance of auditory-visual interaction on auditory localization.

In this study, we seek to expand on the field by having testers take on tests in progressively more complex VEs and analyze how efficient the auditory and visual cues are when used in VEs of varying complexity.

1.2 Objective

The objective of this study is to gain an understanding on how efficient different localization cues can be in helping a user identify the position of a object in a 3D-scene.

Task 1: Run a test in which players are tasked with finding an object in a 3D-environment through auditory cues. This test will take place in three different versions, where every version has its own level of difficulty. (See ”Method” for details on variation in difficulty)
Task 2: Run a test in which players are tasked with finding an object in a 3D-environment through visual cues. This test will take place in three different versions, where every version has its own level of difficulty. (See ”Method” for details on variation in difficulty)

Task 3: Run a test in which players are tasked with finding an object in a 3D-environment through a combined use of visual and audio cues. This test will take place in three different versions, where every version has its own level of difficulty. (See ”Method” for details on variation in difficulty)

Task 4: Analyze the data from these tests and draw conclusions.

1.3 Research questions and hypothesis

Can a difference in efficiency between different localization cues across different audio-visual tests in a VE be measured? If so, how much more efficient is that particular technique compared to another technique when used in VEs of different complexity levels?

The hypothesis is that in an experiment where a test-player is placed in a VE and tasked with finding a specific objects position, the combined use of audio and visual signals can maximize the players efficiency in finding a specific object. Auditory tests will most likely show that it is significantly less efficient than visual signals in communicating the position of an objects position. This relationship is expected to stay the same the same throughout the test. However, the actual disparity between the different techniques as they are used in VEs of different complexity is unknown, and will be determined by the end of this study.
Chapter 2

Chapter content:

2.1 Experiment software description
2.2 Threats to validity
2.3 Preparation and setup
2.4 Participants
2.5 Procedure

In order to accomplish the objective stated in chapter 1.2, the study makes use of an experimental design. This chapter goes into detail on how the experiment was designed. The decisions regarding design choices are also motivated.

2.1 Experiment software description

To determine the efficiency of audio-visual cues in a VE, this study made use of an Experimental design in which several testplayers took on a number of tests in a VE created using the Unity game engine. These tests made use of Visual, Auditory, and the combined Visual+Auditory cues to help testplayers identify objects in their VE. During this test, data regarding how quickly a testplayer managed to find the correct object in the VE, as well as the number of times the testplayer failed to do this, was saved for further analysis.

By monitoring test-players performance the results could be compared to the hypothesis and predictions. The test consisted of a VE in which the test-players were tasked with identifying a random object in the scene. This test could be one out of three stages; one which used audio cues, one which used visual cues, and one which combined both of these. All tests were split up into three difficulty levels in order to gauge how the difficulty between different techniques scaled as tests became more difficult.

Audio: The test uses a first-person camera. The player may not move his character around, but may move his crosshair around. The testplayer is in the middle of a 3D-room, where he or she then was tasked with clicking an object that was making a continuos noise. Upon clicking this object, it stopped making the noise and, after a
brief pause, another object started making a noise. During this test, a timer counted how long it took the player to find each object, in addition to the players accuracy (Was the right object clicked? If not, which one did he click, and at what time?). This data was saved to a file for later analysis.

In order to accommodate an environment in which that the player may properly identify the origin of the sound, a pair of headphones was supplied at the time of the audio test.

Figure 2.1: An illustration of the audio test. The target object in the scene (In this picture, the right-most cube) starts giving off an audio stimulus, which causes the testplayers headphones to give off audio stimulus from the corresponding direction.
**Visual:** The test used the same camera perspective, controls and environment as the audio test. When the test begins, a 3D-arrow appears on the players HUD. This arrow is constantly pointing towards a random object in the scene. The players goal is to click the object that the 3D-arrow is pointing towards. Clicking the right object with the crosshairs will remove the 3D-arrow, which will then reappear after a brief pause, pointing towards another object. The player is tasked with repeating this action a set number of times. During this test, a timer counted how long it took the player to find each object, in addition to the players accuracy (Was the right object clicked? If not, which one did he click, and at what time?). This data was saved to a file for later analysis.

![Figure 2.2](image)

**Figure 2.2:** The visual test. The upper half of the HUD contains a slowly rotating 3D-arrow which will always remain on the same position on the screen and point towards the target object. In the picture, the 3D-arrow is pointing towards the right object. In order to draw attention to the 3D-arrow, color contrast [7] was used to make it stand out from the VE.

**Audio + Visual:** The test used the same camera perspective, controls and environment as the other tests. In this test, the player was tasked with finding objects through the use of both visual and auditory signals. The target object would therefore both emit sound and have an appropriate 3D-arrow pointing towards it on the test-players HUD. The player was tasked with clicking the appropriate object with his crosshair. Upon doing this, the object stopped emitting sound and the 3D-arrow disappeared. After a brief pause, another target would start emitting sound and have a 3D-arrow pointing towards it. During this test, a timer counted how
long it took the player to find each object, in addition to the players accuracy (Was the right object clicked? If not, which one did he click, and at what time?). This data was saved to a file for later analysis.

**Figure 2.3:** An illustration of the Audio+Visual test. The target object is both giving off audio and has a 3D-arrow pointing towards it.

As Wadstein shows in his study on influencing localization behavior [4] artistic choices in geometry, lighting and texturing may influence player decisions in a three-dimensional environment. In order to combat this, the visual and audio+visual test does not use any visual variation on the target models so that player attention would not be drawn away from the target object in the test.
Studies have found that when users are presented with visual and auditory stimuli, the object is identified the fastest when the sound it makes matches the visual image presented [6]. Therefore, to avoid making the testplayers distracted by a sound that would be inappropriate in the context of the visual image presented, a neutral nondescript sound similar to white noise was employed.

In order to gauge how efficient different localization cues are in more complex VEs, all test-variants are divided into three "complexity levels". These add more potential targets to the environment for the player to click, increasing the difficulty of the test. To tell them apart, they were kept to a scale where "1" means the easiest and "3" means most difficult.

Figure 2.4: Complexity level 1 (Easiest VE). Audio 1, Visual 1 and AudioVisual 1 used this VE.
Figure 2.5: Complexity level 2 (Normal VE). Audio 2, Visual 2, and AudioVisual 2 used this VE.

Figure 2.6: Complexity level 3 (Hardest VE). Audio 3, Visual 3 and AudioVisual 3 used this VE.

Figures 2.3-2.6 specify which test used what complexity level. Comparisons of data from these tests can be seen in chapter 3.
2.2 Threats to validity

Object order

Originally, the program was designed so the players would have to click all the object in the scene once in a random order. However, this design had an inherent flaw; For every object the player correctly identified, was one object that no longer could be a target. This meant that players who had clicked all the objects in the scene but one knew exactly which object was going to be the next target; The one object that had not yet been a target. Preliminary tests revealed this flaw, and the program was modified so the player had to click all the objects in the scene once in a random order, apart from the final object which would be randomly decided.

Learning

Preliminary testing revealed another flaw in the software. As the order of the tests was random, different testplayers would go through learning processes at different rates. For example, a testplayer whose first test was the hardest version of the visual test, would have far worse times at this particular test than another testplayer who's first test was the easiest version of the visual test.

As the research question facilitated the need for several difficulty setting of the tests, and a fixed order of tests would cause testplayers to perform better as time went on (Thereby putting the validity of earlier tests at risk) the decision was made to run all of the tests twice. This would remedy the learning process, as all testplayers would be equally prepared to handle the tests the second time around, regardless of test order in the first round.
2.3 Preparation and setup

A small room at Blekinge institute of technology was arranged for this experiment. This room required key-card access, which ensured that testing sessions would remain undisturbed. In order to create a uniform lighting environment for the testers, all drapes were closed and the lamps were turned on. The headphones used in the test were a pair of Panasonic RP-HTX7. The same room and equipment was used for all test players.

![Image of test environment]

**Figure 2.7:** Picture of the test environment. This picture shows both the position of the testplayer (The chair on the left) and the position of the experiment supervisor (ES) during the test (The chair on the right).

In order to gain accurate data for the tests, testplayers had to remain vigilant and focused for the duration of the tests. If a test should take too long, the test-players focus may drop and negatively impact their performance [9]. To combat this, the tests themselves were kept short enough to gather relevant information from the users without making the test redundant.
2.4 Participants

A total of 10 participants were gathered for this test. All participants read and approved of the user consent document (See appendix A). The age range of the participants spanned ages 19 to 25. All testers were male. According to the questionnaire filled out by the testers, all of them regularly played computer games.

2.5 Procedure

The experiment session was set up in the following order:

- User consent
- Pre-questionnaire
- Test information
- Randomize test order and run application, gather data
- Randomize the test order again and run application, gather data
- Post-test questionnaire
**User consent:** Before participating, the users had to read and sign a consent document. This document informed the participants that there was no financial compensation involved in the test and that their participation in the experiment was completely voluntary. It also stated that personal information as well as data collected during the experiment would be kept anonymous. This document was also signed by the ES. A translated version of this document is available in appendix A.1.

**Pre-questionnaire:** After reading and signing the user consent, participants were asked to fill out a pre-questionnaire that asked them about their age, computer game habits, and if they had any handicaps that might influence their performance during the test. A translated version of this document is available in appendix A.2.

**Test information:** The users read a page containing rules for the tests. This involved what their objective in the tests were, how to interact with their VE, and what happened when they were done. A translated version of this document is available in appendix A.3.

**Randomize test order and run application, gather data:** The next part involved both the participant and the ES. The ES sat in front of the laptop (Figure 2.4, the chair on the right) while the tester sat in front of computer display (Figure 2.5). The ES created a random order of the different tests using a random list generator [11], and from the laptop shown in Figure 2.4 initiated the tests in the order of this list. At this point, the tester, sitting in front of the display shown in Figure 2.5, commenced the tests. Upon completing each test, the ES took the data file generated from the program and put it in a folder for later analysis. After doing this, the ES started the next test in the list. This process repeated until all tests had been done in order of the generated list.

**Randomize the test order again and run application, gather data:** After completing the first round of tests, the ES once again created a random order of tests, and had the test-player take them again. Once again, data was collected from all tests.

**Post-test questionnaire:** After completing the second round of testing, the test-players were asked to fill out a form where they were asked which of the test-variants they perceived as being the easiest to complete. A translated version of this document is available in appendix A.4.
In this chapter, the data gathered from the test will be presented. Interpretations and conclusions from this data will then be presented in chapter 4.

### 3.1 Test data

When we compare the overall data from the two rounds of testing (Figure 3.1 and 3.2), we find that the average time to complete each test was significantly faster during the second round. A paired t-test [10] of these times gives us a P-value of 0.0480. As P < 0.05, we may reject the null hypothesis that the tests would result in equal performance. As such, we can conclude that our most accurate data is found in the second round of tests, as these present data from users who are equally familiar with the tests, thereby passing the learning threshold which may have negatively impacted results during the first round (See chapter 2.2 Threats to validity).

As far as efficiency between the different techniques go, the numbers presented in Figure 3.1 and 3.2 shows us which of the tests were finished the fastest. In order to determine which testvariant was faster, tests of corresponding complexity in their VEs were compared to each other. Results of two-tailed T-tests done on these values can be seen in Figure 3.3.
| Tester 1 | 81.66 Errors: 4 | 10.66 Errors: 0 | 7.89 Errors: 0 |
| Tester 2 | 21.46 Errors: 4 | 12.07 Errors: 3 | 3.19 Errors: 0 |
| Tester 3 | 15.67 Errors: 10 | 4.37 Errors: 0 | 2.96 Errors: 0 |
| Tester 4 | 11.86 Errors: 1 | 9.03 Errors: 0 | 4.98 Errors: 0 |
| Tester 5 | 36.11 Errors: 8 | 10.85 Errors: 2 | 14.47 Errors: 1 |
| Tester 6 | 76.935 Errors: 20 | 21.64 Errors: 7 | 30.48 Errors: 2 |
| Tester 7 | 20.516 Errors: 2 | 47.68 Errors: 6 | 3.53 Errors: 0 |
| Tester 8 | 35.515 Errors: 2 | 6.18 Errors: 0 | 3.03 Errors: 0 |
| Tester 9 | 23.551 Errors: 5 | 7.03 Errors: 0 | 10.57 Errors: 1 |
| Tester 10 | 82.735 Errors: 5 | 13.88 Errors: 1 | 9.45 Errors: 1 |

Average time = 40,603 seconds
Average time per cube = 5,8004 seconds
Number of errors = 61

| Tester 1 | 17.84 Errors: 1 | 5.42 Errors: 0 | 3.01 Errors: 0 |
| Tester 2 | 8.9 Errors: 1 | 7.93 Errors: 2 | 2.66 Errors: 0 |
| Tester 3 | 7.53 Errors: 2 | 3.86 Errors: 0 | 2.42 Errors: 0 |
| Tester 4 | 15.58 Errors: 0 | 7.95 Errors: 0 | 3.96 Errors: 0 |
| Tester 5 | 22.064 Errors: 5 | 6.41 Errors: 0 | 2.53 Errors: 0 |
| Tester 6 | 9.14 Errors: 1 | 7.16 Errors: 1 | 2.7 Errors: 0 |
| Tester 7 | 7.08 Errors: 1 | 4.07 Errors: 1 | 3.62 Errors: 0 |
| Tester 8 | 8.93 Errors: 1 | 4.41 Errors: 1 | 2.19 Errors: 0 |
| Tester 9 | 19.55 Errors: 1 | 13.2 Errors: 0 | 2.73 Errors: 0 |
| Tester 10 | 14.22 Errors: 0 | 5.12 Errors: 0 | 3 Errors: 0 |

Average time = 13,0834 seconds
Average time per cube = 1,869 seconds
Number of errors = 13

| Tester 1 | 11.21 Errors: 0 | 23.13 Errors: 0 | 4,35 Errors: 0 |
| Tester 2 | 8.01 Errors: 0 | 4.27 Errors: 0 | 3.44 Errors: 0 |
| Tester 3 | 8.73 Errors: 2 | 3.57 Errors: 0 | 1.67 Errors: 0 |
| Tester 4 | 16.82 Errors: 2 | 5.73 Errors: 0 | 8.69 Errors: 0 |
| Tester 5 | 22.452 Errors: 1 | 10.37 Errors: 1 | 4.48 Errors: 0 |
| Tester 6 | 8.26 Errors: 1 | 6.74 Errors: 1 | 2.42 Errors: 0 |
| Tester 7 | 14.33 Errors: 4 | 4.88 Errors: 2 | 1.94 Errors: 0 |
| Tester 8 | 7.74 Errors: 1 | 5.2 Errors: 0 | 2.58 Errors: 0 |
| Tester 9 | 14.68 Errors: 3 | 8.21 Errors: 0 | 2.82 Errors: 0 |
| Tester 10 | 15.18 Errors: 0 | 4.89 Errors: 0 | 3.64 Errors: 0 |

Average time = 12,7412 seconds
Average time per cube = 1,8201 seconds
Number of errors = 14

Figure 3.1: Table with results for all tests from the first round of testing.
**Figure 3.2:** Table with results for all tests from the second round of testing.

<table>
<thead>
<tr>
<th>Tester 1</th>
<th>Round 2, Audio 3</th>
<th>Errors: 2</th>
<th>Round 2, Audio 2</th>
<th>Errors: 1</th>
<th>Round 2, Audio 1</th>
<th>Errors: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester 2</td>
<td>16,15</td>
<td>9,71</td>
<td>5,841</td>
<td>4,7</td>
<td>2,13</td>
<td>0</td>
</tr>
<tr>
<td>Tester 3</td>
<td>27,73</td>
<td>5,709</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
<tr>
<td>Tester 4</td>
<td>6,126</td>
<td>4,18</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
<tr>
<td>Tester 5</td>
<td>20,24</td>
<td>7,68</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
<tr>
<td>Tester 6</td>
<td>23,31</td>
<td>8,994</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
<tr>
<td>Tester 7</td>
<td>24,25</td>
<td>21,29</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
<tr>
<td>Tester 8</td>
<td>10,516</td>
<td>17,18</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
<tr>
<td>Tester 9</td>
<td>16,57</td>
<td>4,6</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>19,7</td>
<td>9,18</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>32,79</td>
<td>6,55</td>
<td>2,6</td>
<td>4,7</td>
<td>3,24</td>
<td>0</td>
</tr>
</tbody>
</table>

Average time = 19,7382 seconds
Average time per cube = 2,819 seconds
Number of errors = 42

<table>
<thead>
<tr>
<th>Tester 1</th>
<th>Round 2, Visual 3</th>
<th>Errors: 1</th>
<th>Round 2, Visual 2</th>
<th>Errors: 1</th>
<th>Round 2, Visual 1</th>
<th>Errors: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester 2</td>
<td>9,03</td>
<td>4,75</td>
<td>2,81</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 3</td>
<td>9,83</td>
<td>5,42</td>
<td>2,89</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 4</td>
<td>7,87</td>
<td>3,58</td>
<td>2,57</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 5</td>
<td>8,06</td>
<td>5,99</td>
<td>2,89</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 6</td>
<td>12,03</td>
<td>4,64</td>
<td>2,57</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 7</td>
<td>6,16</td>
<td>4,53</td>
<td>2,57</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 8</td>
<td>6,73</td>
<td>3,15</td>
<td>2,03</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 9</td>
<td>7,14</td>
<td>4,746</td>
<td>2,03</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>10,18</td>
<td>6,89</td>
<td>2,03</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>13,36</td>
<td>3,9</td>
<td>2,03</td>
<td>0</td>
<td>2,03</td>
<td>0</td>
</tr>
</tbody>
</table>

Average time = 9,039 seconds
Average time per cube = 1,291 seconds
Number of errors = 11

<table>
<thead>
<tr>
<th>Tester 1</th>
<th>Round 2, Audiovisual 3</th>
<th>Errors: 0</th>
<th>Round 2, Audiovisual 2</th>
<th>Errors: 0</th>
<th>Round 2, Audiovisual 1</th>
<th>Errors: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester 2</td>
<td>10,632</td>
<td>7,31</td>
<td>3,62</td>
<td>0</td>
<td>2,48</td>
<td>0</td>
</tr>
<tr>
<td>Tester 3</td>
<td>12,21</td>
<td>4,13</td>
<td>2,48</td>
<td>0</td>
<td>2,48</td>
<td>0</td>
</tr>
<tr>
<td>Tester 4</td>
<td>6,659</td>
<td>3,13</td>
<td>2,086</td>
<td>0</td>
<td>2,086</td>
<td>0</td>
</tr>
<tr>
<td>Tester 5</td>
<td>7,54</td>
<td>5,95</td>
<td>3,33</td>
<td>0</td>
<td>3,33</td>
<td>0</td>
</tr>
<tr>
<td>Tester 6</td>
<td>21,79</td>
<td>5,45</td>
<td>2,14</td>
<td>0</td>
<td>2,14</td>
<td>0</td>
</tr>
<tr>
<td>Tester 7</td>
<td>8,06</td>
<td>4,38</td>
<td>2,75</td>
<td>0</td>
<td>2,75</td>
<td>0</td>
</tr>
<tr>
<td>Tester 8</td>
<td>5,27</td>
<td>3,06</td>
<td>1,74</td>
<td>0</td>
<td>1,74</td>
<td>0</td>
</tr>
<tr>
<td>Tester 9</td>
<td>5,881</td>
<td>3,98</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>11,8</td>
<td>6,28</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>9,33</td>
<td>4,5</td>
<td>2,54</td>
<td>0</td>
<td>2,54</td>
<td>0</td>
</tr>
</tbody>
</table>

Average time = 9,9272 seconds
Average time per cube = 1,418 seconds
Number of errors = 7

<table>
<thead>
<tr>
<th>Tester 1</th>
<th>Round 2, Audiovisual 3</th>
<th>Errors: 0</th>
<th>Round 2, Audiovisual 2</th>
<th>Errors: 0</th>
<th>Round 2, Audiovisual 1</th>
<th>Errors: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester 2</td>
<td>10,632</td>
<td>7,31</td>
<td>3,62</td>
<td>0</td>
<td>2,48</td>
<td>0</td>
</tr>
<tr>
<td>Tester 3</td>
<td>12,21</td>
<td>4,13</td>
<td>2,48</td>
<td>0</td>
<td>2,48</td>
<td>0</td>
</tr>
<tr>
<td>Tester 4</td>
<td>6,659</td>
<td>3,13</td>
<td>2,086</td>
<td>0</td>
<td>2,086</td>
<td>0</td>
</tr>
<tr>
<td>Tester 5</td>
<td>7,54</td>
<td>5,95</td>
<td>3,33</td>
<td>0</td>
<td>3,33</td>
<td>0</td>
</tr>
<tr>
<td>Tester 6</td>
<td>21,79</td>
<td>5,45</td>
<td>2,14</td>
<td>0</td>
<td>2,14</td>
<td>0</td>
</tr>
<tr>
<td>Tester 7</td>
<td>8,06</td>
<td>4,38</td>
<td>2,75</td>
<td>0</td>
<td>2,75</td>
<td>0</td>
</tr>
<tr>
<td>Tester 8</td>
<td>5,27</td>
<td>3,06</td>
<td>1,74</td>
<td>0</td>
<td>1,74</td>
<td>0</td>
</tr>
<tr>
<td>Tester 9</td>
<td>5,881</td>
<td>3,98</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>11,8</td>
<td>6,28</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>9,33</td>
<td>4,5</td>
<td>2,54</td>
<td>0</td>
<td>2,54</td>
<td>0</td>
</tr>
</tbody>
</table>

Average time = 9,9272 seconds
Average time per cube = 1,418 seconds
Number of errors = 7

<table>
<thead>
<tr>
<th>Tester 1</th>
<th>Round 2, Audiovisual 3</th>
<th>Errors: 0</th>
<th>Round 2, Audiovisual 2</th>
<th>Errors: 0</th>
<th>Round 2, Audiovisual 1</th>
<th>Errors: 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tester 2</td>
<td>10,632</td>
<td>7,31</td>
<td>3,62</td>
<td>0</td>
<td>2,48</td>
<td>0</td>
</tr>
<tr>
<td>Tester 3</td>
<td>12,21</td>
<td>4,13</td>
<td>2,48</td>
<td>0</td>
<td>2,48</td>
<td>0</td>
</tr>
<tr>
<td>Tester 4</td>
<td>6,659</td>
<td>3,13</td>
<td>2,086</td>
<td>0</td>
<td>2,086</td>
<td>0</td>
</tr>
<tr>
<td>Tester 5</td>
<td>7,54</td>
<td>5,95</td>
<td>3,33</td>
<td>0</td>
<td>3,33</td>
<td>0</td>
</tr>
<tr>
<td>Tester 6</td>
<td>21,79</td>
<td>5,45</td>
<td>2,14</td>
<td>0</td>
<td>2,14</td>
<td>0</td>
</tr>
<tr>
<td>Tester 7</td>
<td>8,06</td>
<td>4,38</td>
<td>2,75</td>
<td>0</td>
<td>2,75</td>
<td>0</td>
</tr>
<tr>
<td>Tester 8</td>
<td>5,27</td>
<td>3,06</td>
<td>1,74</td>
<td>0</td>
<td>1,74</td>
<td>0</td>
</tr>
<tr>
<td>Tester 9</td>
<td>5,881</td>
<td>3,98</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>11,8</td>
<td>6,28</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Tester 10</td>
<td>9,33</td>
<td>4,5</td>
<td>2,54</td>
<td>0</td>
<td>2,54</td>
<td>0</td>
</tr>
</tbody>
</table>

Average time = 9,9272 seconds
Average time per cube = 1,418 seconds
Number of errors = 7
**Figure 3.3:** Tables with paired T-test values for all tests. The left table uses times from the first round of testing, and the right table uses numbers from the second round of testing. Numbers in bold represent a P-value that is smaller than 0.05, meaning there is a statistically significant difference between two tests. (Note: the disparity in results between round 1 and round 2 is attributed to the learning effect explained in chapter 2.2. A specific example of how the learning process affected the results can be seen in Round 1 – Visual 2 versus Audio 2. The paired p-value of this particular case is very different not only from the corresponding result in Round 2, but also from the results in Visual 1 versus Audio 1 as well as Visual 3 versus Audio 3, in that it is significantly higher than both of these. Similar cases may be seen in other cases of round 1. The erratic p-values presented from round 1 validates the decision to split up the test in two rounds in order to counterbalance the learning process.)
3.2 Difficulty as complexity of VE increases

Average time spent on tests
(Second round of testing)

Figure 3.4: Average times for each test.

As the complexity of the scene increases, we can see a significant increase in time taken across all tests. Figure 3.4 presents the average times for each test.

Figure 3.5: This table displays the average time to click a cube (in seconds) as well as the increase in time taken to click a cube compared to a less-difficult VE. (Numbers taken from round 2 of tests)

As seen in Figure 3.5, the largest increase in time taken per test, on average, belongs to the audio tests, with audio-visual coming in second and visual being first.

<table>
<thead>
<tr>
<th>Test</th>
<th>Average time taken per cube (Seconds)</th>
<th>Increase in time taken from Test 1 to Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Round 1</td>
<td>Round 2</td>
</tr>
<tr>
<td>Audio 1</td>
<td>3,018</td>
<td>1,47536</td>
</tr>
<tr>
<td>Audio 2</td>
<td>2,86</td>
<td>1,901</td>
</tr>
<tr>
<td>Audio 3</td>
<td>5,8004</td>
<td>2,819</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>Average time taken on tests (Seconds)</th>
<th>Increase in time taken from Test 1 to Test 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Round 1</td>
<td>Round 2</td>
</tr>
<tr>
<td>Audio 1</td>
<td>9,055</td>
<td>4,4261</td>
</tr>
<tr>
<td>Audio 2</td>
<td>14,341</td>
<td>9,5073</td>
</tr>
<tr>
<td>Audio 3</td>
<td>40,603</td>
<td>19,7382</td>
</tr>
</tbody>
</table>
3.3 Perceived difficulty

In addition, the testers were asked to fill out a form after completing both test rounds. This form asked them "Which of the test variants did you perceive as the easiest to clear?" (See Figure A.5 for the original form, translated into English) and gave them the option to answer "The Visual test", "The Audio-based test", or "The Visual + Audio-based test". The result of this may be seen in Figure 3.6 below.

![Table showing the results](image)

**Figure 3.6:** Results from the form that all testers were asked to answer after the tests had been completed.

3.4 Discussion

The results show us that in a virtual 3D environment, the slowest localization stimulus is audio, with visual and audio-visual stimulus being tied for first place. This holds true on the condition that the player is familiar with all of the stimuli to begin with; However, even if the players aren't familiar with the stimuli, a similar if not exactly alike relationship can be found by looking at the results from round one of testing. The results of this experiment may have been different if the experiment allowed the player to move around with conventional first-person-shooter controls as opposed to the static position employed in the experiment. If the players were to use these controls to navigate a more maze-like VE while using the allotted stimuli to find specific targets in it, it is possible we would have gotten different results.
The perceived difficulty was at odds with the gathered data. All testplayers reported
that they found the audiovisual test to be the easiest test to clear. While we may not
be able to reject the null hypothesis of visual versus audiovisual, the average times
we gained from these display slightly faster results for the visual tests than the
audioVisual. In addition to this, Figure 3.1 and 3.2 show that audiovisual tests had,
on average, the fewest amount of errors (i.e clicking the wrong target) among all
test methods. This suggests that the combined stimuli of auditory and visual signals
in the audiovisual test leads to slightly slower times, but less prone to errors than
than visual stimuli alone.
Chapter 4

Conclusions

Chapter content:

4.1 Evaluating the results
4.2 Future work

This chapter will conclude and discuss the results in chapter 3, as well as comparing them to the hypothesis of this study. In addition, the null hypothesis for the paired T-test is discussed and some ideas for further research are presented.

4.1 Evaluating the results

In our hypothesis, we predicted that the different methods would rank accordingly:

- Audio (Slowest method)
- Visual (By default, in between slowest and fastest)
- Audio-visual (Fastest method)

The audio-test was predicted to be significantly less efficient than the visual tests. Audio-visual was predicted to be the most efficient (Due to the combined stimuli) and visual would by default stay between these two.

However, as shown in Chapter 3, Figure 3.2 and Figure 3.3, the actual results do not match for all predictions in the hypothesis. Our paired T-test on all three levels of visual and audio-visual tests cannot reject the null hypothesis. As the null hypothesis states that there is no difference in speed between visual and audio-visual tests, it is therefore inconclusive on the matter of which is the faster method. The audio tests, when compared to audio-visual and visual tests, give us significant results that reject the null hypothesis, and therefore matches the predictions.
The hypothesis suggested that the relationship between the techniques in VEs of varying complexity would be determined at the end of this study. With the results from Figure 3.4 and Figure 3.5, we can see that the complexity of the scenes have a direct influence on how quickly users finish the tests. This shows that tests which feature more complex VEs take, on average, more time per cube. This result also matches the prediction that the relationship between times taken would be stay the same throughout the test.

From this, we can conclude the following order for efficiency among our localization stimuli, under the condition that the player is familiar with each of the stimuli:

- Slowest: Audio
- Fastest: Visual/Audiovisual (Shared, as it is inconclusive which of these two is the fastest method)

### 4.2 Future work

All testplayers reported that they regularly play computer games, which may have influenced the result. This study could benefit from a more diverse amount of participants, involving those who do not play computer games and compare how these perform to those who are more accustomed to computer games. In addition, adding more test subjects could influence the results. With more test-players involved, the paired T-test may need to use a lower threshold to determine statistically significant values. The study may yield additional data if eye-tracking technology would be added to help monitor the players attention in their VE. As results suggest auditory and visual stimuli is a slower but more accurate method than purely visual, it could be interesting to conduct further research between these methods. Research in the automobile industry that study haptic stimuli, heads up and auditory displays may also be of interest in taking this study further. As this study has stated that there is a difference in efficiency between different localization cues, expanding the test with more complex VEs and more subtle localization aids would be interesting. An example of this could be changing the color or lightning of the objects in the scene to draw player attention to a specific object. As this study did not study haptic stimuli, creating a similar test that takes this into account could yield very interesting results.
Appendix A.1

Identification efficiency of objects in 3D-environments through the use of 3D-sound and 3D-arrows

Information to participants and approval of participation

Read this document carefully before you decide to partake in this experiment

Experiment background: This experiment is part of a study which goal is to research efficiency of 3D-audio and visual aids in three-dimensional virtual environments.

Experiment procedure: You will be in G303, a room in BTH. In this room you will use a computer to complete a number of tests in which you interact with a 3D-environment.

Before the experiment begins we ask you to fill out a form in which you give information on you gender and age, as well as information regarding your vision, hearing, and if you regularly play computer games. You have the right to refuse answering any of these questions should you not wish to answer them. (See page 3 for form)

Experiment description: See page 4 in this document.

Expected time for experiment: The test will take roughly 10 minutes to complete.

Compensation: There is no monetary compensation for your participation.

Confidentiality information: Your identity will be kept confidential. This document, if it signed by you, will be kept separate from the data gathered during the experiment. Information collected during the experiment will not be possible to link with your name. The individual results gathered during the experiment will be put in anonymous files, and analysis and conclusions will be based of off these files. Your name will not be used in any report or article.

Voluntary participation: Your participation in this study is voluntary. There is no penalty for not participating.

The right to withdraw from study: You have the right to withdraw from the study at any time without consequence.

Lead Researcher: If you have any questions regarding the study you may contact the Lead Researcher:

Karl Åbom, student in Technical artist for games
Blekinge institute of technology
E-mail: kalleabom@hotmail.com
Appendix A.1

• I am 18 years or older and am competent to provide consent.
• I have read, or had read to me, this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction. The description of the study conducted have been understood by me.
• I agree that my data is used for scientific purposes. I approve that my data is published in scientific publications in a way that does not reveal my identity.
• I understand that my participation in this study is voluntary.
• I understand that I have the right to refuse answering any question I am asked, and that I may withdraw from participation at any time.
• I understand my participation is fully anonymous.

Participant's name:

Participant's Signature:

Date:

Statement of investigator's responsibility: I have explained the nature and purpose of this study as well as which procedures it will involve. I have offered to answer any questions surrounding the study, and I have fully answered these. I am under the impression that the participant understands my explanation and has freely given informed consent.

Investigators signature:

Date:

Disclaimer: Appendix A.1 was originally in Swedish. It has been translated for this thesis.
Appendix A.2

Questionnaire 1

Please fill out/circle the appropriate answer

Question 1  **Gender:**  Male  Female

Question 2  **Age:** _____

Question 3  **How often do you play computer games in your free time?**

- Not at all
- Around 2 hours/week
- Around 5 hours/week
- Around 8 hours/week
- Around 10 hours/week
- Around 15 hours/week
- More than 15 hours a week

Question 4  **Do any of the games that you play fall under the genre of First-person-shooter?**
(For example: Call of Duty, Left 4 dead, Portal) *If you answered "Not at all" on question 3, you may disregard this question and proceed to question 5.*

- Yes
- No

Question 5  **Vision:**  Normal  Glasses  Lenses

Question 6  **Do you have any other eye deficiencies?**


Question 7  **Do you have any problems with your hearing?**

- Yes
- No

Disclaimer: Appendix A.2 was originally in Swedish. It has been translated for this thesis.
Appendix A.3

Description of experiment

You will go through 18 tests. The test environment is a 3D-room with a number of cubes in it. Each test has got the same goal: Use the stimuli available in the test to identify which of the cubes in the room you should click. When you think you have identified this cube, drag your mouse cursor over it and click it using your left mouse button. If you have clicked the correct cube the stimuli will disappear, and after a second has passed, it will reappear on another cube in the scene. The tests are split up in the following varieties:

The Visual test: During this test a rotating three-dimensional arrow will be pointing towards the cube you are supposed to click. Examine the arrow to see where it is pointing to, and click the cube it is pointing towards. If you click the wrong cube nothing will happen; the 3D-arrow will keep pointing towards its cube. If you click the correct cube the 3D-arrow will disappear, and after a second has passed, it will reappear and it will be pointing towards another cube in the scene. Repeat this process until the test automatically turns itself off.

The Audio test: During this test the cube you are supposed to click will be emitting a white noise. Use your hearing to identify which of the cubes the sound is coming from, and click that cube. If you click the wrong cube nothing will happen; the white noise will continue. If you click the correct cube the white noise will stop, and after a second has passed, it will start emitting again from another cube in the scene. Repeat this process until the test automatically turns itself off.

Visual+Audio test: During this test the cube you are supposed to click will be emitting a white noise in addition to having a three-dimensional arrow pointing towards it. Use these stimuli to identify which of the cubes you are supposed to click. If you click the wrong cube nothing will happen; the white noise will continue and the 3D-arrow will keep pointing towards the cube. If you click the correct cube the white noise will stop and the 3D-arrow will disappear, and after a second has passed, the audio and 3D-arrow will return on another cube in the scene. Repeat this process until the test automatically turns itself off.

If there is anything that is unclear, please ask your questions before the test begins.

Disclaimer: Appendix A.3 was originally in Swedish. It has been translated for this thesis.
Questionnaire 2 (To be filled out after you have completed all tests)

Which of the test variants did you perceive as the easiest to clear? (Circle your answer)

The Visual test

The Audio-based test

The Visual+Audio-based test

Disclaimer: Appendix A.4 was originally in Swedish. It has been translated for this thesis.
References


[2] Luca Chittaro and Stefano Burigat, 3D location-pointing as a navigation aid in Virtual Environments, Proceedings of the working conference on Advanced visual interfaces (AVI '04), Figure 1, page 3, New York, NY, USA, 267-274, 2004, ACM.


