

# Assessing a Multimodal User Interface in a Target Acquisition Task

Gustavo Roveló

Dpto. de Sistemas Informáticos y Computación  
Universitat Politècnica de València  
Camino de Vera S/N 46022 Valencia  
España

*gusrorui@posgrado.upv.es*

M.-Carmen Juan

Inst. U. de Automática e Informática Industrial  
Universitat Politècnica de València  
Camino de Vera S/N 46022 Valencia  
España

*mcarmen@dsic.upv.es*

Francisco Abad

Inst. U. de Automática e Informática Industrial  
Universitat Politècnica de València  
Camino de Vera S/N 46022 Valencia  
España

*fjabad@dsic.upv.es*

Emilio Camahort

Inst. U. de Automática e Informática Industrial  
Universitat Politècnica de València  
Camino de Vera S/N 46022 Valencia  
España

*camahort@dsic.upv.es*

**The goal of multimodal user interfaces is to assist users in completing tasks quicker or with fewer errors, by reducing their cognitive workload. Visual, auditory and tactile feedback modalities are commonly used as a means to provide users with information through different sensory channels. Understanding how visual, auditory and tactile feedback modalities complement each other is still an open research area for understanding how computers can help users to complete everyday tasks more efficiently.**

**In this paper, we evaluate user performance on a target acquisition task that simulates a real life scenario: locating a book in a bookcase. We performed a comprehensive comparison among the seven possible combinations of three feedback channels (visual, auditory and tactile). Twenty-four volunteers participated in our experiment.**

**Our results show that using the combination of the three feedback modalities improves user performance on this evidently visual task. However, using the visual modality alone does not increase user performance as much as the other feedback channels. By means of a post-experiment questionnaire, subjects reported that they perceived tactile modality as the most helpful feedback channel.**

*User Performance Evaluation, Multimodal Feedback, Target Acquisition*

©The Authors 2012.

Published by the British Computer Society Ltd.

Proceedings of 26th British HCI Group Annual  
Conference on People and Computers.

©The original version of this article is published in  
the eWiC series at:

[http://ewic.bcs.org/upload/pdf/ewic\\_hci12\\_full\\_paper18.pdf](http://ewic.bcs.org/upload/pdf/ewic_hci12_full_paper18.pdf)

## 1. INTRODUCTION

We perceive the world with our five senses, receiving and processing information from different sources to do our everyday tasks. Using different sensory channels at the same time allows us to improve performance on solving a problem or performing several tasks simultaneously. For example, we are able to process spoken directions given by a GPS device while driving. However, there are some situations that demand complete attention of one of our senses to accomplish the task successfully.

Regarding the driving example, setting a new destination on the GPS requires exclusive attention of the visual sense. Whereas this could be done while driving, user performance decreases, which in this particular case, could increase the probability of a road accident. Multimodal user interfaces address these kind of problems by taking advantage of human's inherent multisensory capabilities.

According to Wickens' Multiple Resource Theory (MRT), a computer system that presents information using multiple sensory channels (i.e. visual, tactile and auditory), is more effective than a system that uses only one modality (Wickens 2002). This is especially true when performing high workload tasks. As we use different senses to perceive complementary information, our brain uses multiple cognitive resources to process all the information, thereby dividing the cognitive workload and increasing user performance. Furthermore according to Wickens' MRT and results from previous research, providing the user with extra information in an already saturated channel is counterproductive (Maltz and Shinar 2007; Wickens 2002; Wickens et al. 2002).

We present the results of an experiment to assess user performance in a target acquisition task. We used visual, tactile and auditory cues (and their combinations) to inform the users on the location of the target. This was through on an everyday task: searching for a book in an unorganized bookshelf. This is predominantly a visual task, in which the user is focused on reading the books' titles while searching for a given book. In our experiment, subjects used a handheld device that can produce three types of feedback. Our goal was to find the feedback channel (or a combination of them) that best improves user performance for this particular task. User performance is measured as the time a user needs to find a given book. We measured the performance of 24 volunteers on this task and, additionally, we asked them to answer a questionnaire about their experience during the experiment.

The outcome of the statistical analysis shows that adding any type of feedback reduces the time required to find a book. As expected, using exclusively the visual channel provides the worst results. Surprisingly, for the other modalities (including the combination of several feedback channels), user performance is very similar. Users reported in the subjective questionnaire that tactile feedback was the most useful and the visual feedback was the least helpful for this task.

Despite our results depending on the application and context, we believe that our findings can be extrapolated to similar environments. For example, applications for helping users to find products in department stores or on warehouse shelves could benefit with the knowledge of which feedback channel helps the user to perform more efficiently. Assistive technologies for visually or hearing impaired persons could be much more effective if we could understand the intricacies of different senses working on a common task.

The rest of the paper is organized as follows. First, we survey related previous research in Section 2. Section 3 describes the features of the application we used in our tests, the task performed by the subjects, and the experimental design of our study. Then in Sections 4 and 5 we present and discuss the outcome of the statistical analysis. Finally, we draw our conclusions and directions for future research in Section 6.

## 2. RELATED RESEARCH

The use of visual, auditory and tactile feedback channels has previously been applied in alerting systems. Other examples of multimodal systems help users in orienting tasks and provide users with other types of information. In this section we review the most relevant papers that set the background for our research.

### 2.1. Alerting Cues

Sun *et al.* studied how visual, tactile and auditory modalities (and their combinations) affect user performance in a steering task (Sun et al. 2010). The user was asked to follow a track with a stylus in the screen of a tablet computer. The system used different feedback channels to alert the user when the stylus was leaving the track. They found that tactile feedback made the best improvement in users' accuracy. However, the completion time of

the task, did not depend on the type of stimuli used. Users rated the combination of auditory and visual cues as the best option. They also pointed out that the vibrations on the stylus made them commit unintended errors.

Visual and auditory feedback modalities were also studied in a collision avoidance system in the work by Maltz and Shinar (2007). Both feedback modalities were tested (individually and combined) to assess how they could help drivers to keep a safe distance from the car in front. They found that using alerting messages improves user performance while driving. Auditory alerts obtained better results than visual and visual and auditory combined, although, users found combined alerts to be more helpful.

Lee and Spence (2008) also evaluated the use of multimodal feedback to assist drivers. They assessed user performance avoiding collisions while completing different tasks on a mobile phone (dialling a number, writing a text message, etc.) The results show that using visual, auditory and tactile cues together as feedback from the mobile device improves drivers' performance. Using more than one feedback channel allowed the user to reduce the amount of attention required to use the phone.

Smith *et al.* explored the differences between visual, tactile and auditory stimuli to help the user in a task-switching situation (Smith *et al.* 2009). Visual and aural are the preferred methods to let users know that they have to direct their attention to the other task. The problem is that these cues may go unnoticed in a visually or sound saturated ambient. Their results show that tactile feedback is not only able to alert the user about a new task, but it also can help the user in locating said task. The tactile-orienting cues performed better than the tactile-alerting cues, but there were no differences with respect to auditory cues.

Another study presented a system that helped older adults with some degree of visual impairment, performing a drag and drop task (Jacko *et al.* 2003). The goal of the experiment was to assess which feedback modality helped the user to drag and drop an icon into a given folder. The system provided feedback when the icon was at the correct screen position. The results show that using any extra feedback as confirmation improves user performance. Nevertheless, colour enhancement alone as visual feedback displayed poor results compared with the rest of the feedback modalities. According to their findings and the results of previous studies, using additional feedback is important for people with reduced visual acuity, but it also represents an advantage for people with normal vision.

Vitense *et al.* also presented a comparison between visual, auditory and tactile feedback as alerting cues for a drag and drop task (Vitense *et al.* 2002). In this case, users with normal vision received feedback when the dragged icon was on top of any folder, not just the correct one. According to their results, auditory and tactile feedback increased task workload compared to visual modality. Visual and tactile feedback combined produced a lower workload similar to that found in visual feedback alone. Trimodal feedback was also found to produce a lower workload. All feedback modalities on their own, as well as auditory plus visual feedback, proved to be helpful in increasing task accuracy. Additionally, tactile feedback, both alone and combined with visual feedback, proved to be a good replacement for visual feedback, reducing the time needed to drop the icon into the target folder.

Finally, visual, auditory and tactile feedback modalities have been researched to improve target selection techniques in 3D scenes (Vanacken *et al.* 2009). In this study auditory and tactile feedback modalities did not improve user performance. The visual modality was sufficient to improve user performance in the selection task in dense 3D scenes where the target object might be occluded.

## **2.2. Orienting Cues and Information Communication**

Bresciani *et al.* (2008) studied how, depending on the type of the activity, background signals in different sensory channels distract users from the activity they are performing. They reported that the visual is the modality most affected by auditory and tactile distracters, and conversely is the one that distracts users' attention the least. The auditory modality was the least susceptible to visual and tactile distracters and the modality that affected the perception of the other two modalities the most. They also found that a background bimodal signal distracted user attention more than a single modal signal.

Wickens *et al.* also studied a system that provides information using both visual and auditory feedback (Wickens *et al.* 2002). They tested each modality and their combination to provide air traffic control information to pilots during a simulated flight. They found that the visual modality affects user performance while air traffic scanning, as the pilots have to deviate their sight to the information display. The auditory feedback worked better for delivering information as short messages.

The tactile feedback channel has been used to give orienting cues, either for improving target finding tasks (Oron-Gilad et al. 2007), or for giving drivers directions and proximity information (Van Erp and Van Veen 2004). The former study compares tactile and visual stimuli to provide the direction and magnitude of the hand movement required to aim at a target. They found that vibrotactile stimulus effectively guides the user's hand, reaching the target faster than when using visual feedback. The latter study compares vibrotactile and visual stimuli to warn the driver of an approaching turn, and how far away it is. They demonstrated that vibrotactile feedback allows drivers to correctly interpret the given information while reducing their visual workload. Nevertheless, despite the fact that visual plus tactile feedback obtained better reaction times, users felt that tactile alone was enough to understand the information clearly and effortlessly.

Ahmaniemi and Lantz (2009) and Charoenchaimonkon *et al.* (2010) also studied different modalities to assist users in a target finding task. The former evaluates the use of tactile feedback to help users find a target in an open space in an AR application. The device vibrates when the user points it towards the target, and the changing frequency of the vibration gives a hint about the distance to the objective. The second paper evaluates the speed and accuracy of users while receiving visual, auditory and tactile feedback. They found that user performance improves more when receiving tactile feedback than when receiving auditory feedback, and this effect is more noticeable when the difficulty of the task increases. They also found that auditory and tactile channels alone work better than using visual feedback in the eminently visual task of target acquisition.

El-Shimy *et al.* propose using aural and tactile modalities in addition to visual feedback for picking targets in a 3D scene rendered on a regular display (El-Shimy et al. 2009). The added modalities are used to provide depth information. The authors evaluated two different ways of providing feedback: discrete feedback, which indicates whether or not the target has been acquired, and continuous feedback, that provides information on target distance. They conclude that discrete feedback improves accuracy of target acquisition. In addition they observed that tactile feedback reduced both reaction time and task completion time but, resulted in a decrease in success rate.

### **2.3. Meta-Analysis**

The following papers present meta-analyses of studies on user performance in applications that provide feedback using different sensory channels.

The first paper presents a comparison between visual feedback, the combination of visual and aural feedback, and the combination of visual and tactile feedback (Burke et al. 2006). Their results show that adding either aural or tactile feedback improved reaction time and user performance, but did not decrease error rate. Furthermore, it seems that tasks under high workload conditions do not benefit from additional aural feedback, since it appears to increase the workload and decrease effectiveness. The meta-analysis revealed that adding tactile feedback improved user performance under the same conditions.

The second meta-analysis compares the combination of visual and tactile feedback with visual feedback alone (Prewett et al. 2006). The authors report that the combination of visual and tactile feedback produces better results than visual feedback alone. This combination reduces reaction times and improves user performance when the workload is high and when multiple tasks are being performed. They also state that their findings show that combining visual and tactile feedback does not affect error rates, and that visual and tactile bimodal feedback have no positive effects on communication tasks. The authors assert that the combination of these modalities produces the best results when providing alerting and orienting cues.

The last meta-analysis also evaluates the use of vibrotactile cues alone and combined with visual feedback (Prewett et al. 2012). As expected, tactile modality alone and combined with visual feedback, improves user performance. The impact of tactile feedback replacing visual feedback depends on task complexity. According to their results, vibrotactile cues adequately replace visual alerts but perform worse when replacing visual direction cues.

All the described research papers perform different studies on user performance for different tasks. They also compare visual, auditory and tactile feedback modalities, or any combination of all three. However, despite the wide variety of studies in the field, there are still many computer-assisted human tasks that have not been evaluated. Certainly, as mentioned by several authors of the reviewed papers, there is a need to perform more studies to understand how human multimodal perception capabilities can be leveraged to better provide task domain information through computer interfaces. Our study describes how to improve user performance on a target acquisition task: finding a book in a bookshelf. We present a comprehensive comparison between visual,

auditory and tactile feedback and all their combinations to direct a users' hand towards the desired book. We believe that the results presented in this paper can contribute to a better understanding of the effect of multimodal feedback on user performance.

### 3. EXPERIMENT

This section describes the design of the testbed used to measure user performance in our task. We also characterise the participants and the task they performed.

#### 3.1. Experimental Setup

Searching for a specific object in a bidimensional arrangement of objects is a common everyday task. Looking for the milk jar in the refrigerator's door or locating our preferred juice brand in the supermarket's shelves are examples of this task. The difficulty of this task depends on the number of shelves, the number of items to choose from, how different they look like, if they are organized in a predetermined, known pattern, etc. This task is especially challenging in unknown environments, even when the items are ordered following some code-based distribution. To study this type of problem, we designed an experiment where users have to find a given book in a bookshelf.

The aim of the computer aid is to improve user performance by means of a handheld device that guides the user's hand to quickly locate the desired book in the bookcase without changing the shelf structure or using cumbersome devices. In our design, the user carries a wireless, compact device that is able to generate the different types of feedback we want to test.

We have implemented a prototype of the device using a Nintendo Wii remote (Wiimote) and a Microsoft Kinect camera. This is just a prototype that fulfills the requirements of our device: it provides the three types of feedback signals, it is not heavy or cumbersome, and has tracking capabilities. An actual implementation of this system could use, for example, a smartphone with a set of markers attached to the shelves to provide tracking. In our application, the user controls the task with the Wiimote's buttons (for example, to ask for the next book title). The Wiimote provides tactile and visual feedback to the user, and the user's hand position is tracked using the Microsoft Kinect camera.

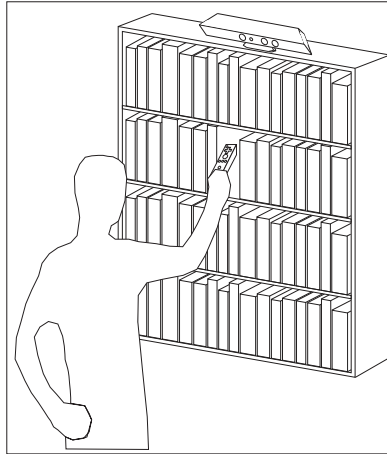
We performed a pilot study to choose among three search strategies: *Find Row First*, *Find Column First* and *Euclidean Distance*. The Find Row First strategy has two steps. In the first, the system helps the user to find the shelf where the book is located. Once the row of books is identified, a radar-like feedback is used in which the stimulus increases as the user's hand gets closer to the desired book. The Find Column First strategy helps the user by guiding her hand to the vertical line where the book is located in the bookshelf, and then increasing the feedback signal as the hand gets closer to the book. The Euclidean Distance strategy increases the stimulus strength as the user's hand gets closer to the book from any direction. We selected the Find Row First strategy to perform our study. In addition to faster searching times, the Find Row First strategy was preferred by most of the users of our pilot study.

#### 3.2. Implementation Details

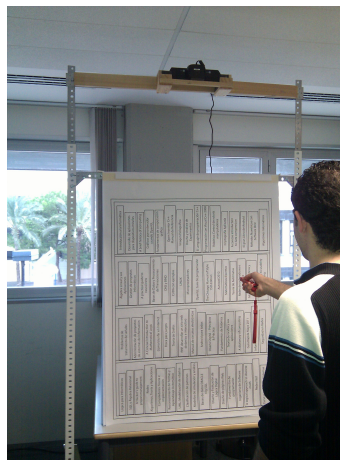
Due to flexibility we designed a model of a bookcase instead of using a real one. Which enabled us to test different book distributions without changing the computer setup. We printed seven book distributions on A0 paper sheets, one per modality (visual, auditory, tactile and their combinations). Each distribution had 65 books in four shelves: 16 books in the top three shelves and 17 books in the last shelf.

Our application uses Microsoft's Kinect to track the user's hand, and a Wiimote as the system interface. The Kinect is placed on top of the bookcase, pointing downwards (see Figure 1). We used the OpenNI library (OpenNI Organization 2010) to connect the Kinect camera to a PC running Windows 7 and to perform user's hand tracking. The application uses a text-to-speech system to broadcast the book title of the target (Clark et al. 2004). Our application communicates wirelessly with the Wiimote using the Wiiuse library (Michael Laforest 2008). Users interact with the system using the Wiimote buttons to start the search task, to listen to the book's title and to confirm the target acquisition (see Figure 2).

The Wiimote provides visual and tactile stimuli. Visual feedback is provided using the Wiimote's LEDs and the vibrator incorporated in this device provides tactile feedback. The Wiiuse library does not support control of the Wiimote speaker, so our system provides auditory feedback using the PC speakers.



**Figure 1:** Schematic representation of our system setup. The Microsoft's Kinect is located on top of the bookshelf model pointing down.



**Figure 2:** The bookcase model used in our tests. The user interacts with the system using Nintendo's Wiimote.

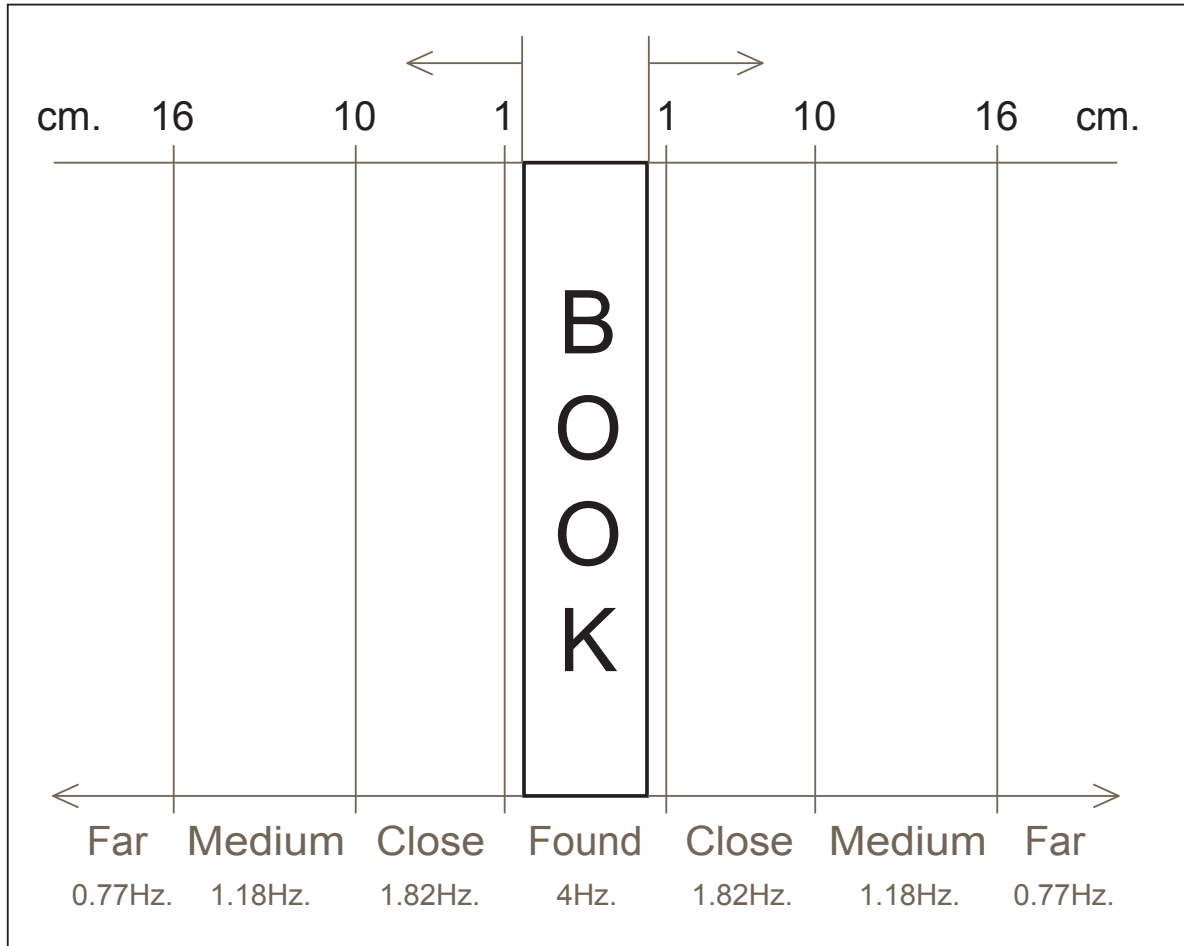
Our system generates four discrete signals to provide information on the distance from the desired book. The four signals are: book found, close to book, medium distance and far from book. When the system provides visual feedback these four signals are generated using the LEDs on the Wiimote. The number of lit LEDs increases as the hand gets closer to the target (one LED means far, four LEDs, book found). For tactile and auditory feedback, the frequency of the pulses changes according to the distance between the Wiimote and the target. The frequencies range from 0.77 Hz. (far) to 4 Hz. (book found). The duration of each pulse is 100 ms. Figure 3 shows the four signal areas, which vary according to the distance from target.

### **3.3. Subjects and Task Description**

Twenty four subjects participated in our experiment, eighteen men and six women. They had a mean age of 27.63 years old (SD= 3.51) ranging from 19 to 34 years old. All the participants were right handed. All subjects performed the task standing in front of the bookcase model, at a distance of one arm. All our subjects were either undergraduate, masters or PhD. students.

In the experiment, each subject had to find twenty four books; three per each of the seven feedback combinations and one extra set for the baseline condition. The baseline condition was searching for a book without assistance. We collected 576 samples in total (72 samples per modality).

Every subject performed the search task starting at the same bookcase model, but using a different feedback modality. Every subject had to search for the same books. We designed the experiment to account for the possible bias when searching for books located in a especially easy or difficult position.



**Figure 3:** We define four discrete areas depending on the distance from target: Found, Close, Medium and Far. The frequency of the tactile and auditory modalities increases as the Wiimote approaches the target. In the visual modality, the number of lit LEDs increases as the users' hand approaches the target.

After completing the task, subjects answered a questionnaire about their experience. On average, subjects needed between 12 and 15 minutes to finish the whole experiment.

### 3.4. Experimental Design

The single factor to account for in our experiment was the feedback modality received by the subject. This factor has eight possible levels (no feedback, and the seven possible combinations of visual, auditory and tactile modalities). We used a within subjects experimental design, the subjects tested all the seven modalities and the baseline condition. We used a Latin Square design to randomly assign the order in which each subject tested each of the feedback modalities.

Before starting the task, each subject received instructions on how to interact with the application. They also performed an initial search to understand how the Find Row First strategy works. Every subject began the experiment searching for three books without any help. This was the baseline condition. Then, the bookshelf model and the feedback modality was changed every three books.

We assessed user performance for each feedback combination measuring the time spent searching for the book ( $T_s$ ). Additionally we evaluated user experience through a questionnaire after the experiment.

## 4. STATISTICAL ANALYSIS

After revising the most recent knowledge and cutting edge innovations described in Section 2, we expect that providing any type of feedback to the user should increase user performance (hypothesis  $H^A$ ). However, as

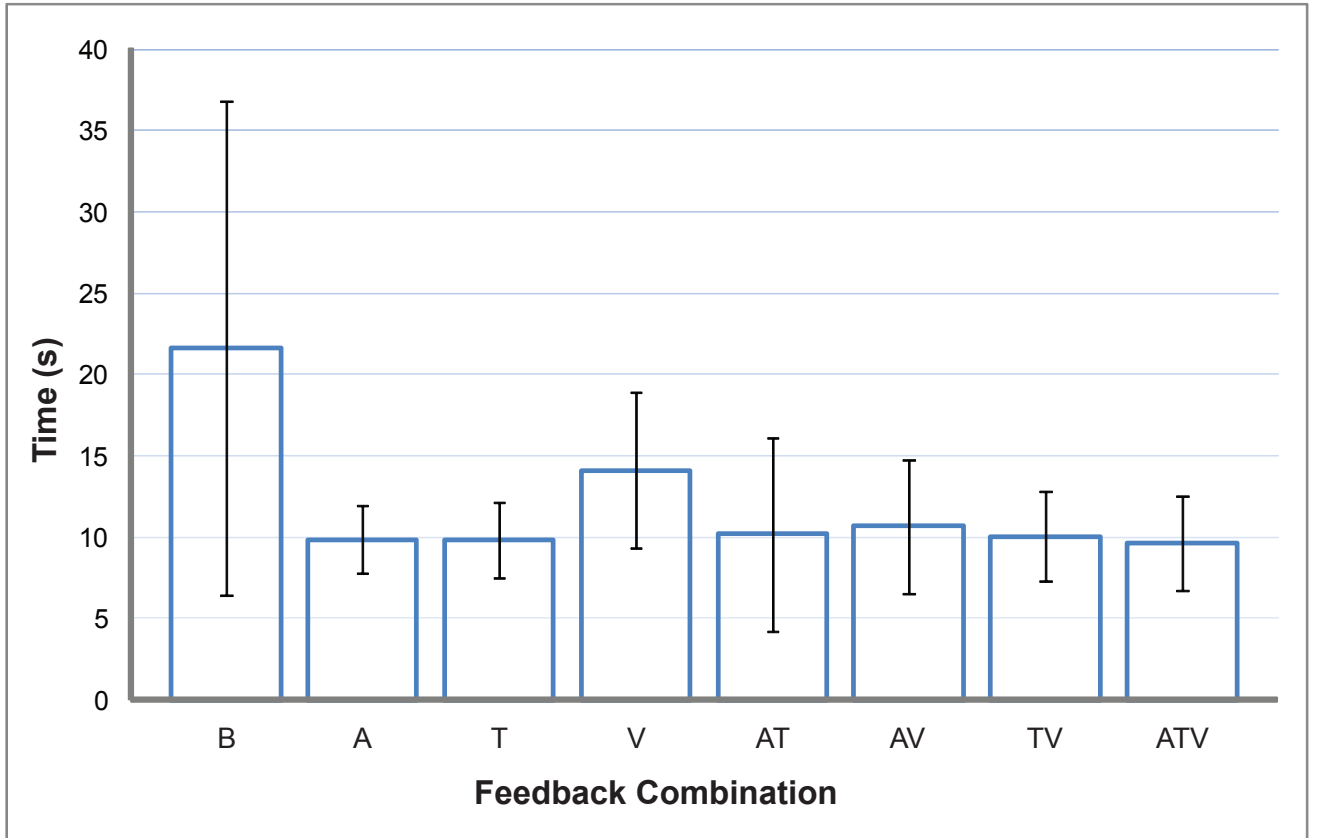
the task is an intensive visual task, we also hypothesize that providing visual feedback should not dramatically improve user performance ( $H^B$ ). Therefore our research hypotheses are:

- $H_0^A$ : providing additional feedback to the user makes no difference in  $T_s$ .
- $H_1^A$ : providing additional feedback to the user affects  $T_s$ .
- $H_0^B$ : providing visual feedback alone makes no difference in  $T_s$  with respect to any other modality.
- $H_1^B$ : providing visual feedback alone is different in  $T_s$  with respect to any other modality.

To study the data captured in the experiment, we performed a One-Way Repeated Measures ANOVA analysis. We selected the Fisher's Least Significant Difference (LSD) method to perform post-hoc tests. All the tests were carried out using a 95 percent confidence level. We used StatGraphics to perform the analysis.

#### 4.1. Results of the Statistical Study

The outcome of the One-way Repeated Measures ANOVA shows a significant effect of feedback modality on  $T_s$  ( $F(7, 575) = 23.53, p < 0.01$ ). The post-hoc test reveals a significant difference between the baseline condition (no feedback) and every other feedback combination tested. It also shows a significant difference between the visual modality alone and all the other modalities, with the visual-only feedback presenting the longest completion time. In contrast, there are no significant differences between the other modalities (including those where visual feedback is present in a bimodal or trimodal combination). Figure 4 shows the average task completion time and standard deviation for each modality. Table 1 reports the average task completion time per subject for each of the seven feedback modalities.



**Figure 4:** Average task completion time and standard deviation per each feedback modality. The bars from left to right represent: Baseline (B), Auditory (A), Tactile (T), Visual (V), Auditory-Tactile (AT), Auditory-Visual (AV), Tactile-Visual (TV) and Auditory-Tactile-Visual (ATV) modalities. The thin lines on each bar represent the standard deviation for each group.

#### 4.2. Questionnaires

We collected subjective information from our volunteers using questionnaires. The questionnaires included yes/no questions, open questions and Likert scale questions. We asked the following questions:



**Table 1:** Average Task Completion Time and Preferred Feedback Channel per Subject.

Subject	A	T	Feedback modality			TV	ATV	Preferred Modality
			V	AT	AV			
1	7.9	11.4	8.2	4.7	10.9	10.3	9.5	A
2	10.5	13.3	16.7	10.7	11.4	15.6	11.4	A
3	8.4	10.1	23.8	11.7	10.3	10.8	9.5	V
4	7.7	11.7	8.6	25.4	8.4	10.5	8.1	T
5	9.1	7.4	8.6	8.1	13.8	9.4	6.2	T
6	7.3	8.2	17.4	12.6	12.3	10.1	10.2	A
7	12.1	9.7	10.9	11.5	11.2	9.4	12.6	T
8	9.2	9.7	6.8	4.6	4.4	6.3	8.7	V
9	9.4	12.3	15.6	5.8	7.3	6.6	9.3	T
10	11.2	8.3	14.6	9.6	13.8	8.9	9.1	T
11	9.6	12.4	27.6	31.2	20.6	9.6	10.8	T
12	10.5	6.3	16.4	6.4	10.6	9.5	7.8	A
13	11.7	8.1	8.9	5.8	5.1	10.8	6.4	T
14	9.1	12.3	12.9	7.6	10.8	9.7	11.6	V
15	13.7	14.5	17.7	5.9	4.7	9.3	8.4	V
16	10.2	8.6	14.5	8.7	7.5	5.8	10.3	A
17	8.5	8.8	18.7	7.7	8.1	10.3	9.1	T
18	8.2	10.7	15.1	10.7	9.5	8.3	8.1	T
19	6.9	6.3	11.9	11.1	16.3	12.2	10.3	T
20	13.2	7.5	12.3	10.4	17.1	15.4	11.8	T
21	14.7	11.5	15.2	9.5	10.3	16.8	20.6	T
22	10.6	6.8	10.9	6.4	6.4	7.5	7.9	T
23	9.7	12.6	13.2	8.7	17.3	11.1	9.2	A
24	6.5	7.6	11.9	9.4	7.4	6.3	4.6	T

- Likert questions.
  - From one to five, how much did you like using the application?
  - From one to five, how easy was using the application?
  - From one to five, how much do you think the application helped you to find the books?
- Yes/no and open questions.
  - Would you use this application in your library?
  - Were you reading the book titles while you were searching for the target?
  - What would you suggest to improve the application?
- Ordering questions.
  - Indicate, in ascending order, the importance of stimuli while using the application.

Regarding how much subjects liked using the application, the majority (66.7 percent) chose the highest value on the scale (five), 29.1 percent liked it (four) and 4.2 percent didn't like it (two in the Likert scale). Fifty percent of the users said that they thought using the system was very easy; 45.8 percent said it was quite easy and 4.17 percent reported that it was neither easy nor complex to use.

Subjects rated the help they thought they received from the device as follows: 50 percent of the subjects reported that it helped them a lot; 45.8 percent stated that it had helped them, and 4.2 percent said the system neither helped nor hindered task execution.

All subjects with the exception of one reported that they would use a system like this in their public libraries. A significant number of users (20.8 percent) did not read the book titles while searching for the target, relying exclusively upon the feedback provided by the system.

Finally, Table 2 shows how subjects ranked the feedback channels (auditory, tactile and visual) with respect to their perceived importance for the task. Table 1 also shows the preferred channel for each subject.

## 5. DISCUSSION

In our experiment, each subject had to find 24 books, consisting of eight rounds of three books in each. In the first round, the user had to search the target without any help from the system. We determined that round as

**Table 2: Feedback Modality Importance**

		Level of importance		
		Highest	Medium	Lowest
Feedback modality	Auditory	25.0%	58.3%	16.7%
	Tactile	58.3%	25.0%	16.7%
	Visual	16.7%	16.7%	66.6%

the baseline condition. In each of the other seven rounds, the user received one of the modalities described in Figure 4. For each user, we measured the time it took to find each of the 24 books.

We can reject our first Null Hypothesis,  $H_0^A$ , as our experiment confirms that receiving any type of feedback allows the user to complete the task faster. The Fisher's LSD post-hoc test revealed statistically significant differences between performing the search process without any help and using any of the studied feedback combinations ( $p < 0.05$ ).

Our second Null Hypothesis,  $H_0^B$  can also be rejected. In our experiment, the average task completion time for subjects receiving visual-only feedback was the highest of all modalities. Despite presenting the slowest completion times, this modality still produced better results when compared with providing no feedback at all (our baseline condition). The Fisher's LSD post-hoc test presented a statistically significant difference between visual-only modality and the remaining modalities ( $p < 0.05$ ). Our results are similar to those reported in the literature. However, we believe this situation might be due to the design of the visual feedback in our application. We intentionally distract users' visual attention away from the bookcase, when looking for visual feedback on the Wiimote. Moving the visual feedback from the Wiimote to the shelves (e.g. using one LED per book in the bookshelf and turning it on to highlight its location), would probably improve user performance using visual feedback. However, this solution, is much more complex to implement in a real environment. The outcome of the statistical analysis did not show any statistical differences between the unimodal, bimodal and trimodal combinations of visual, auditory and tactile feedback (except for the visual-only modality).

Regarding participants' subjective opinion about the experiment, the questionnaires show that 16.7 percent of the subjects expressed that the visual modality was the most helpful. Another 16.7 percent of the users indicated that it was the second best feedback modality, and 66.6 percent reported that it was the least helpful feedback modality. An interesting observation is that, from the four subjects that chose the visual feedback as the best, two obtained the slowest task completion times using this modality alone; one obtained the second slowest time and the other her third quickest average time. Tactile and auditory feedback modalities were regarded as the most helpful (58.3 percent of the subjects chose the vibrotactile cue as the best modality, and 25 percent chose the auditory cue). Users that chose tactile feedback as the best modality stated that vibrations were the most intuitive way to receive information regarding a books' position because they could read books' titles while performing the search task. They also thought that using auditory feedback in a library environment would be annoying for the rest of the users and visual feedback distracted them from the bookshelf. Users that preferred visual feedback mentioned that vibrations made them feel a little uncomfortable and that using LEDs to indicate distance to the book was more intuitive.

Only one subject complained about getting confused when the three modalities were used, as the auditory feedback distracted her from understanding the location feedback. This situation may be related to the findings of Bresciani *et al.*, where subject got confused by the presence of different modalities at the same time (Bresciani *et al.* 2008).

## 6. CONCLUSIONS AND FUTURE WORK

We have studied user performance when using a multimodal device that help users searching for a book in a bookshelf. We measured the task completion time using visual, auditory and tactile cues, alone and combined as feedback modalities to guide users towards the target. Our experiment was performed by 24 volunteers.

Our results show that using any feedback modality improves user performance compared to performing the task without any help. With the exception of visual-only feedback we could not find significant differences between the studied modalities. As expected, adding more signals to an already saturated sensory channel decreased user performance with respect to using other sensory channels.

Subjects completed a questionnaire after the experiment. They reported that the tactile stimulus was more helpful for completing the task than the auditory modality. However, taking into account the results of the statistical analysis, there were no differences between them. Subjects also reported that using visual feedback was the least preferred type of feedback for this task.

We believe that poor user performance using only visual modality can be related to the design of the visual feedback in our application. The subjects had to change their focus from the bookcase to the handheld device to receive the feedback. Embedding visual signals in the shelves would probably provide better results.

As mentioned when we described the experiment, the current implementation of our system would not be practical in large workspaces, due to the high number of tracking devices that would be needed. However, we think that once the effect of the evaluated feedback combination has been understood, a real implementation could design a better suited device, adapted to each specific environment.

Therefore, we plan to evaluate user performance on other types of applications, depending on the seven possible combinations of visual, auditory and tactile feedback. We are especially interested in studying user performance in Augmented Reality applications. Another open issue we think is important to research, is using this kind of applications to help visually impaired people performing everyday tasks such as the one evaluated in this paper.

## ACKNOWLEDGMENTS

This research was supported by the National Council of Science and Technology of México as part of the Special Program of Science and Technology, and by the grant ALFI-3D, TIN2009-14103-C03-03 of the Spanish Ministry of Science and Innovation.

## 7. REFERENCES

- Ahmaniemi, Teemu T., Lantz, Vuokko T. (2009) Augmented Reality Target Finding Based on Tactile Cues. In *Proceedings of the 2009 International Conference on Multimodal Interfaces. ICMI-MLMI '09*, pp. 335–342. ACM, New York, USA.
- Bresciani, J.-P., Dammeier, F., Ernst, M. O. (2008) Tri-modal integration of visual, tactile and auditory signals for the perception of sequences of events. *Brain research bulletin* 75 (6), 753–60. Elsevier, New York, USA.
- Burke, J. L., Prewett, M. S., Gray, A. A., Yang, L., Stilson, F. R. B., Coover, M. D., Elliot, L. R., Redden, E. (2006) Comparing the Effects of Visual-Auditory and Visual-Tactile Feedback on User Performance: A Meta-Analysis. In *Proceedings of the 8th International Conference on Multimodal Interfaces. ICMI '06*, pp. 108–117. ACM, New York, USA.
- Charoenchaimonkon, E., Janecek, P., Dailey, M., Atiwong Suchato, A. (2010) A comparison of audio and tactile displays for non-visual target selection tasks. In *Proceedings of the International Conference on User Science and Engineering. i-USEr, 2010*, pp. 238–243. IEEE, New York, USA.
- Clark, R. A., Richmond, K., King, S. (2004) Festival 2 – build your own general purpose unit selection speech synthesiser. In *Proceedings of the 5th ISCA workshop on speech synthesis*, pp. 173–178.
- El-Shimy, D., Marentakis, G., Cooperstock, J. R. (2009) Tech-note: Multimodal Feedback in 3D Target Acquisition. In *Proceedings of the 2009 IEEE Symposium on 3D User Interfaces. 3DUI '09*, pp. 95–98. IEEE, New York, USA.
- Jacko, J. A., Scott, I. U., Sainfort, F., Barnard, L., Edwards, P. J., Emery, V. K., Kongnakorn, T., Moloney, K. P., Zorich, B. S. (2003) Older adults and visual impairment: what do exposure times and accuracy tell us about performance gains associated with multimodal feedback? In *Proceedings of the SIGCHI conference on Human factors in computing systems. CHI '03*, pp. 33–40. ACM, New York, USA.
- Lee, J.-H., Spence, C. (2008) Assessing the Benefits of Multimodal Feedback on Dual-Task Performance Under Demanding Conditions. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 1. BCS-HCI '08*, pp. 185–192. British Computer Society, Swinton, UK.
- Maltz, M., Shinar, D. (2007) Imperfect In-Vehicle Collision Avoidance Warning Systems Can Aid Distracted Drivers. *Transportation Research Part F: Traffic Psychology and Behaviour* 10 (4), 345–357. Elsevier, New York, USA.
- Michael Laforest (2008) Wiiuse - The Wiimote C Library. <http://sourceforge.net/projects/wiiuse> (23 June 2011).
- OpenNI Organization (2010) OpenNI Library. [www.openni.org](http://www.openni.org) (January 4th., 2012).

- Oron-Gilad, T., Downs, J. L., Gilson, R. D., Hancock, P. A. (2007) Vibrotactile guidance cues for target acquisition. *IEEE Transactions on Systems, Man and Cybernetics. Part C, Applications and Reviews* 37, 993–1004. IEEE, New York, USA.
- Prewett, M., Elliott, L., Walvoord, A., Coover, M. (2012) A Meta-Analysis of Vibrotactile and Visual Information Displays for Improving Task Performance. *IEEE Transactions on Systems, Man, and Cybernetics. Part C: Applications and Reviews* 42 (99), 123–132. IEEE, New York, USA.
- Prewett, M., Yang, L., Stilson, F., Gray, A., Coover, M., Burke, J., Redden, E., Elliot, L. (2006) The benefits of multimodal information: a meta-analysis comparing visual and visual-tactile feedback. In *Proceedings of the 8th International Conference on Multimodal Interfaces*, pp. 333–338. ACM, New York, USA.
- Smith, C., Clegg, B. A., Heggstad, E. D., Hopp-Levine, P. J. (2009) Interruption Management: A Comparison of Auditory and Tactile Cues for Both Alerting and Orienting. *International Journal of Human-Computer Studies* 67 (9), 777–786. Elsevier, New York, USA.
- Sun, M., Ren, X., Cao, X. (2010) Effects of multimodal error feedback on human performance in steering tasks. *Journal of Information Processing* 18, 284–292. KIPS, Seoul, Republic of Korea.
- Van Erp, J. B., Van Veen, H. A. H. C. (2004) Vibrotactile In-Vehicle Navigation System. *Transportation Research Part F: Traffic Psychology and Behaviour* 7 (4-5), 247–256. Elsevier, New York, USA.
- Vanacken, L., Grossman, T., Coninx, K. (2009) Multimodal selection techniques for dense and occluded 3D virtual environments. *International Journal of Human-Computer Studies* 67, 237–255. Elsevier, New York, USA.
- Vitense, H., Jacko, J., Emery, V. (2002) Foundation for improved interaction by individuals with visual impairments through multimodal feedback. *Universal Access in the Information Society* 2, 76–87. Springer Berlin / Heidelberg.
- Wickens, C. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science* 3 (2), 159–177. Taylor & Francis, Oxfordshire, UK.
- ARL-02-8/NASA-02-6 (2002) *Modality Differences in Advanced Cockpit Displays: Comparing Auditory Vision and Redundancy for Navigational Communications and Traffic Awareness*. Wickens, C. D., Goh, J., Helleberg, J., Talleur, D. A. Nasa Ames Research Center.