Visual vs Auditory Augmented Reality for Indoor Guidance

Andrés-Marcelo Calle-Bustos[1], Jaime Juan[1], Francisco Abad[1][a], Paulo Dias[2][b], Magdalena Méndez-López[3][c] and M.-Carmen Juan[1][d]

[1]*Instituto Universitario de Automática e Informática Industrial, Universitat Politècnica de València, 46022 Valencia, Spain*
[2]*Department of Electronics, Telecommunications and Informatics, University of Aveiro, Portugal*
[3]*Departamento de Psicología y Sociología, IIS Aragón, Universidad de Zaragoza, 44003 Teruel, Spain*

Keywords: Augmented Reality, Visual Stimuli, Auditory Stimuli, Indoor Guidance.

Abstract: Indoor navigation systems are not widely used due to the lack of effective indoor tracking technology. Augmented Reality (AR) is a natural medium for presenting information in indoor navigation tools. However, augmenting the environment with visual stimuli may not always be the most appropriate method to guide users, e.g., when they are performing some other visual task or they suffer from visual impairments. This paper presents an AR app to support visual and auditory stimuli that we have developed for indoor guidance. A study (N=20) confirms that the participants reached the target when using two types of stimuli, visual and auditory. The AR visual stimuli outperformed the auditory stimuli in terms of time and overall distance travelled. However, the auditory stimuli forced the participants to pay more attention, and this resulted in better memorization of the route. These performance outcomes were independent of gender and age. Therefore, in addition to being easy to use, auditory stimuli promote route retention and show potential in situations in which vision cannot be used as the primary sensory channel or when spatial memory retention is important. We also found that perceived physical and mental efforts affect the subjective perception about the AR guidance app.

1 INTRODUCTION

Outdoor navigation is already a mature technology with several well-known commercial applications that provide very good mapping and navigation information. Most of these commercial applications are based on Global Positioning Systems (GPS). In comparison, indoor navigation lags far behind since indoor tracking technologies have many limitations and no clear standard has yet emerged. Nevertheless, since mobile devices have evolved and now integrate many technologies and sensors, it is already possible to obtain indoor localization with enough precision to develop systems that can assist users in indoor navigation. Therefore, we are interested in helping users move between different locations in an indoor environment.

In this paper, we develop, explore and evaluate two different guidance systems. The first uses visual stimuli which are displayed on the screen of a mobile device. Stimuli that are overlaid on the camera view using Augmented Reality (AR). The second is based on auditory information that is made available as necessary while the user is moving around the environment. The objective of this work is to evaluate the influence of AR stimuli (visual and auditory) on the global navigation task, as well as on route memorization. A user study was carried out with 20 users to test the developed apps and to evaluate and understand the advantages and limitations of both guidance approaches.

In our study, we consider three hypotheses. The first hypothesis (H1) is that the auditory condition will be effective for indoor guidance. The second hypothesis (H2) is that the auditory condition will require more time than the visual condition. Our

[a] https://orcid.org/0000-0001-5896-8645
[b] https://orcid.org/0000-0002-3754-2749
[c] https://orcid.org/0000-0002-4249-602X
[d] https://orcid.org/0000-0002-8764-1470

Calle-Bustos, A., Juan, J., Abad, F., Dias, P., Méndez-López, M. and Juan, M.
Visual vs Auditory Augmented Reality for Indoor Guidance.
DOI: 10.5220/0010317500850095
In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 1: GRAPP*, pages 85-95
ISBN: 978-989-758-488-6

argument for this second hypothesis is that the auditory condition requires more attention in comparison with the visual condition because of the perceptual dominance of the sense of sight. In other words, since the auditory condition transmits less information, it needs more effort to decode or interpret the signals. Our study involves subjects without visual or hearing impairments, and under 55 years. Therefore, our third hypothesis (H3) is that there will be no statistically significant difference for the performance outcomes due to age and gender.

The paper is organized as follows. After the introduction, we introduce some related work in Section 2. In Section 3, we present the main characteristics of the AR guidance app developed, and in Section 4, we present the user study. Finally, the discussion and our conclusions are presented in Sections 5 and 6, respectively.

2 RELATED WORK

Outdoor navigation problems have gradually been solved over the years, while indoor navigation still has many issues that require further attention (Vanclooster et al., 2016). An Indoor Positioning System (IPS) is able to determine the position of an object or a person in a physical space continuously in real time (Gu et al., 2009). Indoor navigation requires a much more precise tracking technology than GPS (the standard for outdoor navigation). However, no low-cost technology that is precise enough has yet emerged, so there are few indoor navigation systems available.

Indoor navigation presents many additional challenges when compared to outdoor navigation. It needs to be more precise and convey routing information more efficiently than outdoor routing due to the difference of scale. Therefore, AR, especially AR based on SLAM (Simultaneous Localization and Mapping), can be a natural response for IPS given its ability to superimpose virtual content (namely routing and directions) on a view of the real-world environment.

Most AR systems are inherently related to sight, since it is our predominant sense. However, there are several works that use other senses such as hearing. Ribeiro et al. (Ribeiro et al., 2012) created a natural user interface using sound. They presented a spatialized (3D) audio synthesizer to place virtual acoustic objects at specific coordinates of the real world, without explicitly telling the users the locations of objects. The utility of sound in AR environments has not been studied extensively. A few previous works have already shown the following: participants using spatialized sound perform more efficiently and faster than working without sound (Rumiński, 2015); 3D sound in AR environments significantly improves the performance of the task and the accuracy of depth judgment (Zhou et al., 2004); and 3D sounds contribute significantly to the sense of presence and collaboration (Zhou et al., 2004).

On the other hand, very few studies have compared visual and auditory stimuli. One of these is the study presented by Cidota et al. (Cidota et al., 2016). They carried out a study to compare the effects of visual and audio notifications on workspace awareness using AR in a remote collaboration scenario. The participants received AR assistance to solve a physical 2D assembly puzzle called Katamino. Their study (N=12) showed that, regardless of the difficulty level of the task, users preferred visual notifications to audio notifications or no notification. In our study, visual and auditory stimuli are also compared, but for indoor guidance.

2.1 Augmented Reality

Rehman & Cao (Rehman & Cao, 2017) presented an AR-based indoor navigation system. The environment was scanned, and its visual features (3D point clouds) were stored as trackables. Locations and navigation-related information were then associated with those trackables. The 3D point clouds and device orientation were tracked using the camera and inertial sensors of the device. They carried out a study (N=39) of navigation tasks to compare the performance of the participants using a wearable device (Google Glass), an Android Cell Phone (Samsung Galaxy S4), and a paper map. Their results indicated that the performance using the paper map was worse in terms of taking longer and having higher workloads than the two digital navigation tools. However, the performance using the digital tools was worse for route memorization.

Polvi et al. (Polvi et al., 2016) presented a 3D positioning method for SLAM-based handheld AR (SlidAR). SlidAR uses epipolar geometry and 3D ray-casting for positioning virtual objects. They carried out a study involving 23 participants. They compared the SlidAR method with a device-centric positioning method. Their results indicated that SlidAR required less device movement, was faster, and received higher scores from the participants. SlidAR also offered higher positioning accuracy. Piao & Kim (Piao & Kim, 2017) developed an adaptive monocular visual–inertial SLAM for real-time AR apps in mobile devices. Egodamage & Tuceryan (Egodagamage & Tuceryan, 2018) developed a collaborative AR framework based on distributed monocular visual SLAM.

Chung et al. (Chung et al., 2016) developed a projection-based AR guidance system, which projected guidance information directly on the real world. This system was compared with a mobile screen-based guidance app, which shows information on the screen (N=60). The main conclusion was that navigation with the projection-based AR was more natural.

AR has also been used as a planning aid for navigation in a public transportation network. Peleg et al. (Peleg-Adler et al., 2018) developed a route-planning task for public transportation and studied the effects of aging (N=44). They compared the performance of younger and older participants using a mobile AR app and a non-AR app on a mobile phone. The mobile AR app augmented a wall map with the bus schedule for each station. The participants using the AR app completed the task in less time, regardless their age, but with higher error rates when compared to the non-AR app. Chu et al. (Chu et al., 2017) designed mobile navigation services with AR. A study (N = 49) comparing the performance of the participants using AR and maps showed that AR navigation was faster for finding the correct location.

Some AR car navigation systems have also been published (Akaho et al., 2012). Wintersberger et al. (Wintersberger et al., 2019) studied the effect of using AR aids reflected on a windshield on the acceptance and trust of drivers. The environment was a country road with dense fog. In their study (N=26), they used TAM (the Technology Acceptance Model) and an adapted version of the Trust in Automation Questionnaire as measures. Their results showed that augmenting traffic objects that are relevant for driving can increase users' trust as well as other acceptance parameters.

2.2 Audio Guidance

The two most closely related works to ours are the ones by Lock et al. (Lock et al., 2017) and Yoon et al. (Yoon et al., 2019). Lock et al. (Lock et al., 2017) presented ActiVis, which is a multimodal navigation system that was developed using Tango SDK (the same SDK that we have used). Their system guided the user toward a location target using voice commands and spatialized sound. Vibration was used to avoid obstacles. However, the system was not tested with users.

Yoon et al. (Yoon et al., 2019) presented Clew, which is an iOS app that was developed using ARKit. The routes have to be recorded previously with a smartphone and loaded afterwards to provide guidance along the route. Clew included sound, speech, and haptic feedback. The authors highlight two use cases. The first one allows the user to record a route in an indoor environment and then navigate the route back to the initial location. The second use case allows the user to record a route in an indoor environment, store this route, and then navigate the route forward or backward at a later time. Clew can be downloaded from the App Store. Based on their study, the authors concluded the following: ARKit is robust enough for different indoor navigation environments; the motion estimation of ARKit is accurate for navigation routes of around 60 meters; and routes shorter than 33 meters are rated positively by users. The main difference between Clew and our proposal is that our app works reliably on routes of more than 60 meters.

Katz et al. (Katz et al., 2012) presented an AR guidance system for visually impaired users (NAVIG). NAVIG combined input data provided through satellite-based geolocation and an ultra-fast image recognition system. Guidance was provided through spatialized semantic audio rendering.

3 INDOOR GUIDANCE APP WITH AR VISUAL AND AUDITORY FEEDBACK

In this work, a SLAM-based AR app for indoor guidance has been developed. The app supports two different types of indications for indoor routing: visual signs and audio clips. The modules necessary to support guidance features were developed using the Google Tango motion-tracking SDK, which allows mobile devices (equipped with the appropriate hardware) to track their position and orientation throughout 3D space. To increase the reliability of the tracking, the area learning feature of Google Tango, which allows the device to locate itself in a previously known environment, was used. The app was also developed using the Unity game engine and C# scripts.

3.1 Functionality of the App

The developed app requires a configuration step. This configuration step is the same for visual or auditory navigation. First, the supervisor must explore the environment in order to create the required area description file. In the second step, the supervisor defines the possible virtual paths in the environment. Virtual paths are made of connected cells. The steps to be performed by the supervisor are: 1) set a path seed at a given location; 2) move to the desired target position (cells are added next to the previous cell in order to create a line of cells) (Figure 1a); and 3) anchor the last cell as the target position. The process can be repeated to define multiple paths between different locations. After the configuration, the navigation routing app allows users to explore the environment while providing routing indications. The app uses the current user's location and the path information defined in the configuration step to compute the best route to the location target.

Depending on the navigation mode (visual or auditory), the app shows the appropriate visual clues using arrows on the floor (Figure 1b) or plays auditory clips (e. g., "Turn left", "Turn right", "Go forward", or "Stop"). Depending on the condition used, a different method is used to indicate the users to go back to the correct route when they deviate from the path. With visual feedback, a 3D object (Figure 1c) appears showing the position the user should return to. With auditory feedback, given the absence of visual feedback, an additional audio clip is triggered to convey the "out of path" message.

The app also supports rerouting in cases where the user follows an alternative route, either by ignoring the route information or due to the appearance of an obstacle during navigation in the real environment. The app detects these situations, recomputes, and updates the shortest path to the final target.

As stated above, two navigation modes are available.

Visual Navigation. This module presents visual arrows indicating the route to follow (Figure 1b), and a location icon appears when deviating from the route (Figure 1c). When the location icon is presented, a message is also displayed indicating to the user that she/he must reach the icon in order to return to the correct route.

Auditory Navigation. The auditory navigation is significantly more complex than the visual one since the app needs to constantly monitor the position and orientation of the user/mobile device in order to provide the appropriate auditory cues. The app uses a series of pre-recorded audio clips that run according to the instructions coming from the navigation app. An additional message is available when the device shakes significantly, instructing the user to stabilize it. This option was added because conflicting audio directions might be provided when the orientation of the device changes rapidly. The participants held the device in their hands so that the cameras on the device can identify the position and orientation of the device relative to the environment. The device screen did not show anything. Audio was played through the device's speakers.

(a)

(b)

(c)

Figure 1: Examples of the AR guidance app with visual stimuli: (a) configuration step; (b) arrows showing the path; (c) location icon to return to the correct route.

3.2 Architecture of the App

The different functions of the app are distributed in four modules that encapsulate certain functionalities. A module is defined by a series of scripts, which are divided into two parts (core and handlers). The core contains an interface file and a module file. The functions defined in the interface are implemented in the module file. Handlers allow higher customization for certain events.

The architecture of our app consists of four modules (Nav, NavCells, NavVision, and NavAudio). NavCells is used in the configuration and allows the placement of a series of cells on which guidance is performed. The other modules Nav, NavVision, and NavAudio carry out the guidance process together. The scenes make use of the modules. A scene can directly access NavCells and Nav modules. The access to NavVision and

NavAudio modules is achieved through the Nav module.

The NavCells Module. A set of characteristic points provided by the Tango Learning Area is obtained after carrying out a scanning session. In order to carry out correct guidance, including avoiding obstacles, it is necessary to indicate which areas of the environment are walkable, which is not possible using only the stored points. This is the main task of the NavCells. In order to specify different paths, the stored environment is loaded, and a series of cells are placed along the route. These cells are the units that together form the route. Figure 2 shows an example of cell distribution to form the paths of a possible environment. The NavCells module offers three functions for placing cells. Based on the current cell (in the center of the screen), the supervisor might: 1) create a row of cells (moving and anchoring successive cells to the previous one); 2) create a single cell anchored to the previous one; and 3) store the position and orientation of the cells in a file.

The Nav Module. The Nav module is responsible for controlling all guidance. Two processes are necessary (calculation and navigation). In the calculation process, the cells are recovered and the navigation calculation is performed, i.e., the cells that must be followed to reach the desired location target. Once the route that the user must follow has been obtained, the navigation process involves the NavVision and NavAudio modules. The Nav module does the route calculations and invokes the functions of the NavVision and NavAudio modules to show users the shortest route to the target.

The public functions, which are visible to the programmer, allow the cells that form the environment map to be entered and the target to be reached by guiding to be selected. Certain activations are also allowed using properties. These properties handle the activation of the three modules involved in guidance (Nav, NavVision, and NavAudio). When the Nav module is activated, the route calculation is performed from the position of the device to the target. To perform this calculation, the different cells are traversed using a width search algorithm, which stops when the target is found. This search can be performed because each cell contains a list of neighboring cells. These lists are computed when the map is created by finding the connections between adjacent cells.

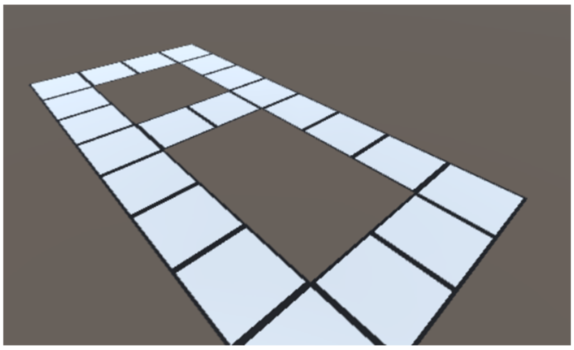

Figure 2: Example of cell distribution to form the paths of a possible environment (image from Unity).

Once the path to a given target is obtained, the navigation mode (visual or auditory) must be activated. In visual mode, the NavVision module is in charge of rendering arrows on the cells to indicate the route. In audio mode, the NavAudio module is used to provide auditory instructions. These two modes can be used together or separately. The visual mode shows the visual elements (arrows). The auditory mode obtains the device orientation in order to check that it is in the right orientation and plays audio clips to give indications to turn, stop, and move forward on the route. Constant checks are carried out between the orientation of the device and the occupied cell to retrieve the direction to follow in order to reach the next cell of the route.

If the user does not follow the indications and leaves the path, a mechanism is triggered to guide the user back to the closest point on the path. This process differs depending on the guidance mode used. For the visual mode, a 3D object appears, which indicates the position to which the user must return. The auditory mode does not show any element, but plays different audio clips to guide the user back to the correct path. Another situation that may occur is the modification of the optimal path. For example, when the user makes a detour on an alternative path. This deviation can occur for different reasons: the user does not follow the appropriate instructions, or the appearance of an obstacle. In these cases, the app recomputes the shortest path to the target.

The NavVision Module. The NavVision module provides visual guidance functionality. This module displays the arrows to indicate the path and displays the redirect icon on the closest location in the route. While the device is on a path, arrows are shown towards the target. When the user is off the path, the location icon appears and a message is displayed on the screen, indicating that the path has been abandoned and the user must search for the location icon to return to the right path. Figures 1b and 1c show the modeling of these objects.

The NavAudio Module. The NavAudio module offers functionalities for guidance using auditory stimuli. Its functionality is similar to that of a music player. The NavAudio module plays, stops, or stores the desired audio clips. The NavAudio module does not perform any calculations. It only responds to the instructions received from the Nav module, which is in charge of performing the relevant calculations. The only mechanism that is activated within the NavAudio module occurs when the device moves abruptly, signaling the user to stop this type of movement and stabilize the device. This module contains the following audios: turn left; turn right; advance; stop; stop shaking; back to route; out of path (the user has left the path, the user must stop to redirect); orienting; reached target; and next target (orienting towards the next target).

4 USER STUDY

A within-subjects user study was conducted in order to first evaluate the two feedback modes of the guidance app. The second goal was to compare the performance and subjective perceptions of users using the two guidance modes. Finally, we also wanted to gather some insight about route memory retention with the two AR stimuli.

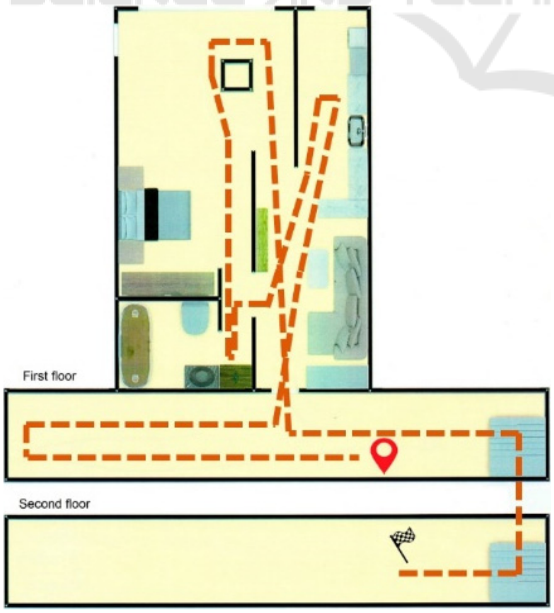

Figure 3: The indoor space map and the route to follow.

4.1 Participants, Procedure, and Measures

A total of 20 participants were involved in the study, of which 12 were women (60%). The age range of the participants was between 7 and 54 years old (35.15 ± 15). The study was approved by the Ethics Committee of the Universitat Politècnica de València, Spain, and was conducted in accordance with the declaration of Helsinki.

All of the participants were involved in two conditions: VisualCondition and AuditoryCondition. The participants were randomly assigned to two groups (Group A and Group B). The participants in Group A completed the navigation task with visual stimuli first and completed the navigation task with auditory stimuli after a period of at least two days. The participants in Group B completed the navigation task using the app with auditory stimuli first and completed the navigation task using the app with visual stimuli after a period of at least two days. Both groups were balanced in such a way that there were 9 users in Group A (45%) and 11 users in Group B (55%). The proportion of women was similar for both groups. The sessions with the users were carried out on two floors of a building with a total space of 60 m^2. Figure 3 shows the apartment map and the route to follow.

The protocol is the following. First, the supervisor configured the environment for this study. The participants were not involved in any way in setting up the environment and did not know the route. Second, the supervisor explained the navigational task to the participant, giving basic explanations about the navigational app. Users of Group A started with the visual stimuli, while users in Group B started with the auditory stimuli. The study was divided into two steps. During Step 1, the supervisor explained globally the overall task to the user, and, using a Lenovo Phab 2 Pro, the user performed the navigation task in the building with either the visual stimuli or the auditory stimuli (depending on the group). The app stored the time required to complete the navigation task as well as the distance travelled. After completion, the user was asked to indicate the route she/he followed on 2D paper maps (Map task). Finally, the user was asked to fill out Questionnaire 1 regarding subjective perceptions. In Step 2, more than two days after the first step, the users were asked to repeat the task with the other stimuli and to fill out of Questionnaire 2 (only to compare the two types of stimuli).

For the Map task, the participants saw two empty maps, corresponding to the first and the second floors of the building. Each map was a simplified two-dimensional map illustrating stairs, hallways, corridors, rooms, and furniture. The two maps were printed on A4-sized paper. The supervisor asked the participants to draw the route they thought they had followed on the maps using a pen. The performance score on the map task (Map) measured the proximity to the route that the users should have followed with values on a scale from 0 to 10. A template with the correct route was used for scoring. Points were subtracted when the route deviated from the correct one. This calculation was done manually. All scores were assigned by the same supervisor to ensure that the criteria were the same for all users.

Questionnaire 1 consists of 16 questions, which are grouped in the following variables: usability, enjoyment, competence, concentration, expertise, calmness, physical effort, mental effort, and satisfaction. Questionnaire 1 was specifically designed for this study and was based on previously used questionnaires (Brooke, 1996; Calle-Bustos et al., 2017; Munoz-Montoya et al., 2019). Questionnaire 2 was designed to evaluate the users' preference regarding the two AR guidance stimuli and consisted of the following questions: 1) Which stimuli did you like the most?; 2) Why?; 3) Which one do you think is the best as a navigational tool?; and 4) Why?.

5 RESULTS

The normality of the data was verified using the Shapiro-Wilk test. The tests indicated that the performance outcomes fit a normal distribution, while the subjective scores did not fit a normal distribution. Therefore, we used parametric tests for the performance outcomes and non-parametric tests for the subjective scores. A statistically significant difference at level $\alpha = .05$ is indicated by the symbol **. The R open source statistical toolkit (https://www.r-project.org) was used to analyze the data (specifically, R version 3.6.2 and RStudio 1.2.5033 for Windows).

To determine whether or not there were order effects for the distance traveled and the time required to complete the navigational task for visual or auditory stimuli, the four possible combinations were analyzed. First, we considered the Distance variable. To determine whether or not there was an order effect for this variable and VisualCondition between the participants who the used visual stimuli first (61.99 ± 5.04) and the participants who used the visual stimuli second (62.86 ± 3.91), we applied the unpaired t-test ($t[18] = -.41$, $p = .685$, $d = .19$). To determine whether or not there was an order effect for the Distance variable and AuditoryCondition between the participants who used the auditory stimuli first (72.80 ± 5.67) and the participants who used the auditory stimuli second (76.42 ± 6.57), we applied the unpaired t-test ($t[18] = -1.25$, $p = .226$, $d = .56$). Second, we considered the Time variable. To determine whether or not there was an order effect for this variable and VisualCondition between the participants who used the visual stimuli first (171.78 ± 36.62) and the participants who used the visual stimuli second (146.67 ± 38.04), we applied the unpaired t-test ($t[18] = 1.42$, $p = .174$, $d = .64$). To determine whether or not there was an order effect for the Time variable and AuditoryCondition between the participants who used the auditory stimuli first (292.64 ± 40.36) and the participants who used the auditory stimuli second (284.44 ± 34.76), we applied the unpaired t-test ($t[18] = .46$, $p = .654$, $d = .21$). These results indicate that there were no statistically significant order effects for the distance traveled and the time required to complete the navigational task for visual or auditory stimuli. Therefore, since there was no order effect, the participants were grouped by condition.

5.1 Performance Outcomes

To determine how the use of visual or auditory stimuli affects the navigation using the app, we compared the performance outcomes between the two conditions (VisualCondition vs. AuditoryCondition) (within-subjects analysis). First, we considered the variable that indicates the total time in seconds used to perform the task (Time). To determine whether or not there were differences for this variable between the conditions of VisualCondition (157.97 ± 39.44) and AuditoryCondition (288.95 ± 38.16), we applied the paired t-test ($t[19] = -14.16$, $p < .001$**, $d = 3.17$). This result indicates that there were significant differences between the two conditions. The participants of the AuditoryCondition spent more time completing the task.

Second, we considered the variable that indicates the total distance in meters traveled by the user to complete the task (Distance). To determine whether or not there were differences for this variable between VisualCondition (62.47 ± 4.48) and AuditoryCondition (74.43 ± 6.35), we applied the paired t-test ($t[19] = -7.94$, $p < .001$**, $d = 1.77$). This result indicates that there were significant differences between the two conditions. The participants of the AuditoryCondition walked longer.

The variable that represents the participants' memory of the route followed in the task was also analyzed (Map). To determine whether there were differences for this variable between the participants who used the visual stimuli first (6.22 ±3.79) and the participants who used the auditory stimuli first (9.36 ±1.15), we applied the unpaired t-test (t[18] = -2.47, p = .024**, d = 1.11). These results indicate that there were significant differences between the two groups in favor of the participants who used the auditory stimuli, who remembered the route followed in the task better.

5.2 Gender and Age Analysis

To determine if gender influences the Distance variables for VisualCondition, we applied the unpaired t-test (*t[18]* = 1.94, p = .068, d = .89); for AuditoryCondition, we also applied the unpaired t-test (*t[18]* = .78, p = .447, d = .35). To take into account gender and the Time variable, the same test was given to VisualCondition (*t[18]* = 1.69, p = .107, d = .77) and to AuditoryCondition (*t[18]* = .98, p = .342, d = .45). No statistically significant differences were found in any of these analyses. Therefore, we can conclude that the performance outcomes were independent of the participants' gender.

To determine if age influences the Distance variable, we applied an ANOVA test to VisualCondition (F[1,18] = 1.181, p =.292) and to AuditoryCondition (F[1,18] = .873, p =.363). Similarly, to take into account age and the Time variable, the same test was given to VisualGroup (F[1,18] = .065, p =.802) and to AuditoryCondition (F[1,18] = .392, p =.539). No statistically significant differences were found in any of these analyses. Therefore, we can conclude that the performance outcomes were independent of the participants' age.

5.3 Subjective Perceptions

The participants' subjective perceptions about the AR app were measured using Questionnaire 1. The questions were grouped in the following variables: usability, enjoyment, competence, concentration, expertise, calmness, physical effort, mental effort, and satisfaction. We applied the Mann-Whitney U test for all of the subjective variables and compared the participants' perceptions by using the visual stimuli first vs the auditory stimuli. The only significant difference found was for the Satisfaction variable (U = 74.5, Z = 2.040, p = .046**, r = .456) in favor of the visual stimuli.

When the visual stimuli was used first and considering gender, no statistically significant differences were found in any of these analyses. When the auditory stimuli was used first and considering gender, significant differences were found for enjoyment (U = 25, Z = 2.128, p = .043**, r = .642); non-mental effort (U = 25, Z = 2.121, p = .044**, r = .640); and satisfaction (U = 28.5, Z = 2.505, p = .016**, r = .755) in favor of the male group.

We used Spearman's correlation to test the associations among the subjective variables. For the participants who used the visual stimuli first, enjoyment correlated with satisfaction (r = .99, p <.001) and correlated marginally with non-physical effort (r = .61, p = .07). Usability correlated with non-mental effort (r = .77, p = .016) and correlated marginally with calmness (r = .63, p = .071). Perceived competence correlated with non-physical effort (r = .85, p =.004). Non-physical effort correlated marginally with satisfaction (r = .62, p = .075). For the participants who used the auditory stimuli first, enjoyment correlated with non-mental effort (r = .96, p <.001) and satisfaction (r = .76, p =.006). Usability correlated marginally with calmness (r = .57, p = .069). Non-physical effort correlated with satisfaction (r = .75, p = .008).

5.4 Preferences and Open Questions

When asked "Which stimuli did you like the most?", most of the participants (60%) preferred the visual stimuli. The arguments for the users who preferred the visual were the following: it was easier (71%); it was more entertaining (13%), it required less attention (8%); and it was more direct (8%). The arguments for the users who preferred the auditory were the following: it was more entertaining (86%); and it required less attention (8%). When asked "Which one do you think is the best as a navigational tool?", most of the participants (85%) preferred the visual stimuli.

6 DISCUSSION

We developed a new AR app for indoor guidance. Our proposal can be replicated using other devices and SDKs, such as the iPad Pro which can be programmed using ARKit. ARKit 3.5 allows programming apps using the new LiDAR scanner and depth-sensing system that are built into the iPad Pro.

Our study compared the performance outcomes and subjective perceptions of the participants using our AR guidance app with visual and auditory stimuli. To our knowledge, no study such as the one presented here has been conducted.

Our results show that there were statistically significant differences regarding the time required for completing the task and the distance traveled in favor of the visual condition. These results are in line with previous research stating that the sight is the dominant sense in humans (Cattaneo et al., 2008; Papadopoulos & Koustriava, 2011).

After analyzing the participants' memory of the route followed in the task, the results showed significant differences between the two groups in favor of the participants who used the auditory stimuli, who remembered the route followed in the task better. Our argument for this result is that the auditory stimuli forced the participants to pay more attention, which led to better memorization of the route followed. On the other hand, the visual guidance partly overlaps the visual information of the route, thus overshadowing the memorization of the details. The participant focuses more on the app itself than on the path being taken (Blanco et al., 2006; Chung et al., 2016). Taking into account the differences, our results are in line with the work of Rehman & Cao (Rehman & Cao, 2017), which compared paper maps with AR apps. They found that the AR apps required less time and had lower workload, but had worse outcomes in memorizing the route. When using paper maps, even though the participants required more time and traveled more distance, they reached the location target, and, moreover, they memorized the route better. Therefore, we can conclude that H1 ("the auditory condition will be effective for indoor guidance") and H2 ("the auditory condition will require more time than the visual condition") have been corroborated.

Our results show that the performance outcomes (Distance and Time variables) were independent of the age and gender of the participants. This demonstrates that, regardless of gender and age, our AR guidance app has proven to be suitable for indoor guidance. These results are in line with previous works such as (Juan et al., 2014) and corroborates H3 ("there will be no statistically significant difference for the performance outcomes due to age and gender").

With regard to the subjective variables, the results indicate that the app was highly appreciated by the participants using both visual and auditory stimuli in all of the variables analyzed. On a scale of 1 to 7, the means were very high, equal to, or above 6 in all cases. The only significant difference found was for the Satisfaction variable in favor of the visual stimuli. Our first argument regarding this result is that in the visual condition less effort is required to pay attention, and therefore the participants are more satisfied with this condition. Our second argument is related to the dominance of visual information in the human being; the participants are more satisfied when visual stimuli are used.

When the visual stimuli were used first, no statistically significant differences were found in any of the subjective variables between men and women. When the auditory stimuli were used first, significant differences were found in favor of the men for enjoyment, less mental effort required, and satisfaction. The auditory condition is more expensive in terms of processing. In general, women tend to report worse navigation or greater difficulty, including their self-perception of how well they did (Mendez-Lopez et al., 2020). Since the auditory condition is more difficult, gender effects are observed in some variables related to the perception of how the participants evaluate themselves after executing the task.

The correlations among the subjective variables for the two types of stimuli were: 1) the more enjoyment experienced, the more satisfaction felt; 2) the higher degree of usability, the calmer the participant; and 3) the less physical effort required, the more satisfaction felt. An additional correlation between two subjective variables for the auditory stimuli was: the more enjoyment experienced, the less mental effort required. Additional correlations among the subjective variables for the visual stimuli were: 1) the more enjoyment experienced, the less physical effort required; 2) the higher degree of usability, the less mental effort required; and 3) the greater the perceived competence, the less physical effort required. From these correlations, we can argue that perceived physical and mental effort considerably affects the subjective perception that the user has about the app.

With regard to the questions "Which stimuli did you like the most?", 60% of the users preferred the visual stimuli. For the question "Which one do you think is the best as a navigational tool?", 85% of the participants preferred the visual stimuli. At this point, we would like to highlight that the participants involved in our study had no vision impairments and that sight was their dominant sense, so they understandably preferred the visual stimuli. Moreover, with the visual stimuli, the participants saw 3D objects that were mixed with the real environment, while with the auditory stimulus, they only heard audio clips.

A limitation of our work is the sample size and the within-subject design. Using a within-subject design, there could be memory effect that might influence the results. Therefore, it would have been more recommendable that the sample would be larger and the study used a between-subject design. However, in this work in order to determine whether or not there were order effects for the performance variables used, the possible combinations were analyzed. The results indicate that there were no statistically significant order effects for the performance variables used for visual or auditory stimuli. Therefore, since there was no order effect, the participants were grouped by condition using a within-subject design.

7 CONCLUSIONS

This paper presents the development of an AR app for indoor guidance. Our AR guidance app works in any indoor environment and can be used in several rooms or on several floors of the same building. The supervisor configures the route information and creates as many paths as desired.

For the first time, we have carried out a study in which visual and auditory stimuli are compared for indoor guidance. From the results, we can conclude that both visual and auditory stimuli can be used for indoor guidance. The auditory condition required more time and more distance to complete the route, but facilitated a better memorization of the route followed. The performance outcomes were independent of gender and age. Therefore, auditory stimuli can be used for indoor guidance, showing potential in situations in which vision cannot be used as the primary feedback channel or when spatial memory retention is important.

As future work, several studies can be conducted, especially those related to the suitability of the AR guidance app for different groups. For example, our guidance app using the auditory stimuli could be validated with participants with vision problems. In another study, using a larger sample, the results of different adult age groups could be compared (young adults, middle-age adults, and old adults). In that study, it could be observed how age and familiarity with technology influence the results. Our app can use visual and auditory stimuli together. Our hypothesis is that an overall increased performance could be achieved using the two stimuli together. A study for checking this hypothesis could be carried out. Our proposal could also be compared with Clew for routes of less than 33 meters (Yoon et al., 2019). It would also be interesting to further investigate how the auditory sense, including spatial sound and other sensory modalities (e.g., vibration to avoid obstacles) could be used for indoor guidance. Another possible work could be to develop an application for the Map task. In this way, the score of the Map task will be objective. Using a tablet and on the touch screen, the participant will draw the route on a digital map. This digital route will be compared with the correct route and thus obtain objective scores.

ACKNOWLEDGEMENTS

We would like to thank all of the people who participated in the study. We would like to thank the reviewers for their valuable suggestions. This work was funded mainly by FEDER/Ministerio de Ciencia e Innovación – Agencia Estatal de Investigación/AR3Senses (TIN2017-87044-R); other support was received from the Gobierno de Aragón (research group S31_20D) and FEDER 2020-2022 “Construyendo Europa desde Aragón”.

REFERENCES

Akaho, K., Nakagawa, T., Yamaguchi, Y., Kawai, K., Kato, H., & Nishida, S. (2012). Route guidance by a car navigation system based on augmented reality. *Electrical Engineering in Japan*, *180*(2), 897–906. https://doi.org/https://doi.org/10.1002/eej.22278

Blanco, M., Biever, W. J., Gallagher, J. P., & Dingus, T. A. (2006). The impact of secondary task cognitive processing demand on driving performance. *Accident Analysis and Prevention*, *38*, 895–906. https://doi.org/10.1016/j.aap.2006.02.015

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*, *189*(194), 4–7.

Calle-Bustos, A.-M., Juan, M.-C., García-García, I., & Abad, F. (2017). An augmented reality game to support therapeutic education for children with diabetes. *PLOS ONE*, *12*(9), e0184645. https://doi.org/10.1371/journal.pone.0184645

Cattaneo, Z., Bhatt, E., Merabet, L. B., Pece, A., & Vecchi, T. (2008). The Influence of Reduced Visual Acuity on Age-Related Decline in Spatial Working Memory: An Investigation. *Aging, Neuropsychology, and Cognition*, *15*(6), 687–702. https://doi.org/10.1080/13825580802036951

Chu, C. H., Wang, S. L., & Tseng, B. C. (2017). Mobile navigation services with augmented reality. *IEEJ Transactions on Electrical and Electronic Engineering*, *12*, S95–S103. https://doi.org/10.1002/tee.22443

Chung, J., Pagnini, F., & Langer, E. (2016). Mindful navigation for pedestrians: Improving engagement with

augmented reality. *Technology in Society*, *45*, 29–33. https://doi.org/10.1016/j.techsoc.2016.02.006
Cidota, M., Lukosch, S., Datcu, D., & Lukosch, H. (2016). Comparing the Effect of Audio and Visual Notifications on Workspace Awareness Using Head-Mounted Displays for Remote Collaboration in Augmented Reality. *Augmented Human Research*, *1*, 1. https://doi.org/10.1007/s41133-016-0003-x
Egodagamage, R., & Tuceryan, M. (2018). Distributed monocular visual SLAM as a basis for a collaborative augmented reality framework. *Computers and Graphics (Pergamon)*, *71*, 113–123. https://doi.org/10.1016/j.cag.2018.01.002
Gu, Y., Lo, A., & Niemegeers, I. (2009). A survey of indoor positioning systems for wireless personal networks. *IEEE Communications Surveys and Tutorials*, *11*(1), 13–32. https://doi.org/10.1109/SURV.2009.090103
Juan, M.-C., Mendez-Lopez, M., Perez-Hernandez, E., & Albiol-Perez, S. (2014). Augmented reality for the assessment of children's spatial memory in real settings. *PLoS ONE*, *9*(12), e113751. https://doi.org/10.1371/journal.pone.0113751
Katz, B. F. G., Kammoun, S., Parseihian, G., Gutierrez, O., Brilhault, A., Auvray, M., Truillet, P., Denis, M., Thorpe, S., & Jouffrais, C. (2012). NAVIG: augmented reality guidance system for the visually impaired. *Virtual Reality*, *16*, 253–269. https://doi.org/10.1007/s10055-012-0213-6
Lock, J., Cielniak, G., & Bellotto, N. (2017). A portable navigation system with an adaptive multimodal interface for the blind. *AAAI Spring Symposium - Technical Report*, 395–400.
Mendez-Lopez, M., Fidalgo, C., Osma, J., & Juan, M.-C. (2020). Wayfinding Strategy and Gender – Testing the Mediating Effects of Wayfinding Experience, Personality and Emotions. *Psychology Research and Behavior Management*, *Volume 13*, 119–131. https://doi.org/10.2147/PRBM.S236735
Munoz-Montoya, F., Juan, M.-C., Mendez-Lopez, M., & Fidalgo, C. (2019). Augmented Reality Based on SLAM to Assess Spatial Short-Term Memory. *IEEE Access*, *7*, 2453–2466. https://doi.org/10.1109/ACCESS.2018.2886627
Papadopoulos, K., & Koustriava, E. (2011). The impact of vision in spatial coding. *Research in Developmental Disabilities*, *32*(6), 2084–2091. https://doi.org/10.1016/j.ridd.2011.07.041
Peleg-Adler, R., Lanir, J., & Korman, M. (2018). The effects of aging on the use of handheld augmented reality in a route planning task. *Computers in Human Behavior*, *81*, 52–62. https://doi.org/10.1016/j.chb.2017.12.003
Piao, J. C., & Kim, S. D. (2017). Adaptive monocular visual-inertial SLAM for real-time augmented reality applications in mobile devices. *Sensors (Switzerland)*, *17*, 2567. https://doi.org/10.3390/s17112567
Polvi, J., Taketomi, T., Yamamoto, G., Dey, A., Sandor, C., & Kato, H. (2016). SlidAR: A 3D positioning method for SLAM-based handheld augmented reality. *Computers and Graphics (Pergamon)*, *55*, 33–43. https://doi.org/10.1016/j.cag.2015.10.013
Rehman, U., & Cao, S. (2017). Augmented-Reality-Based Indoor Navigation: A Comparative Analysis of Handheld Devices Versus Google Glass. *IEEE Transactions on Human-Machine Systems*, *47*(1), 140–151. https://doi.org/10.1109/THMS.2016.2620106
Ribeiro, F., Florencio, D., Chou, P. A., & Zhang, Z. (2012). Auditory augmented reality: Object sonification for the visually impaired. *2012 IEEE 14th International Workshop on Multimedia Signal Processing, MMSP 2012 - Proceedings*, 319–324. https://doi.org/10.1109/MMSP.2012.6343462
Rumiński, D. (2015). An experimental study of spatial sound usefulness in searching and navigating through AR environments. *Virtual Reality*, *19*, 223–233. https://doi.org/10.1007/s10055-015-0274-4
Vanclooster, A., Van de Weghe, N., & De Maeyer, P. (2016). Integrating Indoor and Outdoor Spaces for Pedestrian Navigation Guidance: A Review. *Transactions in GIS*, *20*(4), 491–525. https://doi.org/10.1111/tgis.12178
Wintersberger, P., Frison, A. K., Riener, A., & Von Sawitzky, T. (2019). Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality. *Presence: Teleoperators and Virtual Environments*, *27*(1), 46–62. https://doi.org/10.1162/PRES_a_00320
Yoon, C., Louie, R., Ryan, J., Vu, M. K., Bang, H., Derksen, W., & Ruvolo, P. (2019). Leveraging augmented reality to create apps for people with visual disabilities: A case study in indoor navigation. *ASSETS 2019 - 21st International ACM SIGACCESS Conference on Computers and Accessibility*, 210–221. https://doi.org/10.1145/3308561.3353788
Zhou, Z., Cheok, A. D., Yang, X., & Qiu, Y. (2004). An experimental study on the role of 3D sound in augmented reality environment. *Interacting with Computers*, *16*, 1043–1068. https://doi.org/10.1016/j.intcom.2004.06.016.