

Enhancing Visuospatial Map Learning through Action on Cellphones

MOUNIA ZIAT, Center for Intelligent Machines, McGill University and Northern Michigan University
CARMEN AU, AMIN HAJI ABOLHASSANI, and JAMES J. CLARK, Centre for Intelligent Machines,
McGill University

The visuospatial learning of a map on cellphone displays was examined. The spatial knowledge of human participants was assessed after they had learned the relative positions of London Underground stations on a map via passive, marginally active, or active exploration. Following learning, the participants were required to answer questions in relation to the spatial representation and distribution of the stations on the map. Performances were compared between conditions involving (1) without auditory cues versus continuous auditory cues; (2) without auditory cues versus noncontinuous auditory cues; and (3) continuous auditory cues versus noncontinuous auditory cues. Results showed that the participants performed better following active and marginally-active explorations, as compared to purely passive learning. These results also suggest that under specific conditions (i.e., continuous sound with extremely fast tempo) there is no benefit to spatial abilities from active exploration over passive observation; while continuous sound with moderate to fast tempo is effective for simple actions (i.e., key press).

Categories and Subject Descriptors: H.1.2 [Models and Principles]: User/Machine.Human factors—*Human factors*; H.5.2 [Information interfaces and Presentation]: User Interfaces—*Auditory (nonspeech) feedback; Evaluation / methodology*; I.3.6 [Methodology and Techniques]: Interaction Techniques

General Terms: Design, Experimentation, Human Factors, Performance

Additional Key Words and Phrases: Cell phones, Human attention, multimodal perception, auditory feedback, map exploration, map learning

ACM Reference Format:

Ziat, M., Au, C., Abolhassani, A. H., and Clark, J. J. 2012. Enhancing visuospatial map learning through action on cellphones. *ACM Trans. Appl. Percept.* 9, 1, Article 5 (March 2012), 15 pages.
DOI = 10.1145/2134203.2134208 <http://doi.acm.org/10.1145/2134203.2134208>

1. INTRODUCTION

Humans have for a long time used maps to represent their near and far spaces. The evolution of maps has followed the technologies that have been used to display them. Since the 1950's, computers have provided indispensable support, allowing users to navigate through interactive maps by scrolling or zooming. More recently, computerized map displays have become portable, and hence more useful, allowing people to explore an environment in the same manner as they do when they carry paper maps. As new services are offered in the field of mobile technologies, travelers are more often carrying their mobile phones during vacations and using them for map navigation. However, the usefulness of the

Contact author's email address: M. Ziat, Psychology Department, Northern Michigan University; email: mziat@nmu.edu.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 2 Penn Plaza, Suite 701, New York, NY 10121-0701 USA, fax +1 (212) 869-0481, or permissions@acm.org.

© 2012 ACM 1544-3558/2012/03-ART5 \$10.00

DOI 10.1145/2134203.2134208 <http://doi.acm.org/10.1145/2134203.2134208>

display of maps on such devices is hampered by the small size of screens. Indeed, it is often difficult to identify a specific location on a map on a small screen, and it is even more difficult to link the spatial layout of different landmarks relative to each other or to the observer. Traditional scrolling and panning methods with joysticks, scrollbars or touch screen input offer only limited support [Guiard et al. 1999] in exploring large-scale maps on small displays. This suggests that we need to consider interactions more suited to the small size of the display, such as handwriting, gesture recognition, or speech recognition [Forman and Zahorjan 1994], as well as methods that compensate for the lack of visual information by using other modalities (auditory or tactile).

1.1 Overview and Related Work

The research presented in this article investigates alternative map scrolling modes by considering multiple-modality and its influence on human visuospatial map learning on a cellphone. Several studies showed that participants using mobile maps (by using a handheld GPS unit or a cell phone) have poor spatial knowledge acquisition [Willis et al. 2008; Krüger et al. 2004; Ishikawa et al. 2008], but no study has focused on active and passive exploration on small displays for visuospatial learning. Indeed, in most of the studies on spatial knowledge on a cellphone [Munzer et al. 2006], participants were always in the situation of a passive observation (i.e., no interaction with the device). The screen display was updated in relation to the participants' position, but they did not trigger any action to change the information displayed on the device. Farrell et al. [2003] suggested that passive observation refers to participants who do not make navigational decisions even when they are in motion. Moreover, Ishikawa et al. [2008] showed that the pedestrians made several stops, walked slowly, or opted for longer paths when they are using a GPS. The authors did not discuss explicitly the sensori-motor aspect, but we believe that the sensory-motor coupling was very poor in these situations because there was no association between participants' actions and the instructions displayed on the device. Indeed, GPS systems make the users perform actions, since they have to follow the instructions given by the device (e.g., in 300m turn left, go ahead, etc.), while during a sensori-motor coupling, the process happens at the same time, that is, users perform actions and, according to their sensations, they modulate their actions in order to modulate their sensations again, and so on. There is no such process while using a GPS navigation device, and which probably leads to a poor spatial knowledge. We believe that the main reason is the lack of active exploration. Situations where actions are coherent with the visual cues might compensate for the small size of the screen, and thus improve users' spatial knowledge acquisition. For example, tilting has been suggested for map navigation by several studies [Rekimoto 1996; Hinckley et al. 2000; Oakley et al. 2004]. Chatting [Chatting 2008] evaluated the action of the user (tilt) and the reaction of the display (movement direction of the map) by testing several mappings and showed that participants made fewer errors and had faster time rates when both direction of the tilt and the map are consistent. Oakley et al. [2000] added either a discrete or a continuous vibrotactile feedback associated to the scrolling/tilting speed. They showed that short haptic vibrations were less efficient than the continuous vibrations. While some [Rekimoto 1996; Hinckley et al. 2000] operated the scrolling with a tilt followed by a button press, others [Oakley et al. 2000; Chatting 2008] operated the scrolling by the tilt only. In this perspective, Dong et al. [2005] showed that a tilt method was equal to multifunction buttons for map navigation on a PDA, but the tilt method has the advantage of helping users create better cognitive overviews of the map. However, none of these studies evaluated the effect of spatial layout learning during passive vs. active exploration and the benefits of the tilt over a stationary situation.

Another factor that may help users to learn spatial knowledge on the visual display during exploration tasks is that of additional modalities such as audition. Brewster showed that the use of *earcons*, which are musical tones that represent specific items, as auditory cues helps navigation in hierarchical structures [Brewster 1998]. Darken and Sibert [1993] used acoustic landmarks as

spatial audio cues to give rough direction-finding, and help navigation in virtual environments. In this study the sound was only audible on a specific range around the landmark. Other studies used acoustic beacons, which are sound cues used for navigation tasks, for wayfinding that is either continuous [Dodiya and Alexandrov 2008; Tran et al. 2000] or noncontinuous [Walker and Lindsay 2003; Lokki et al. 2000; Tran et al. 2000]. Tran et al. showed that a continuous versus pulsed beacon sound did not affect the overall localization accuracy, and participants preferred a continuous beacon sound over a pulsed sound [Tran et al. 2000].

1.2 The Role of Action and Multiple Sensory Modalities Map Exploration Using Small Displays

The main focus of this article is to show that correlated motor action and efficient auditory information can improve visuospatial learning and memorizing the relationships between landmarks on a cellphone map. Imagine a person who wants to locate different subway stations around a specific zone but does not have an access to the internet on the phone or to a paper map. One solution is to rely on one of the map displays in the city, but he/she might have to walk several blocks in order to find one. Another solution is to have a map saved on the cell phone and check all the stations surrounding the area. Like a paper map, an interface with efficient multimodal cues should allow the user to check this information once and be able to memorize it without the need to check it again.

Our series of experiments involved requiring participants to learn on a cellphone the relational space of specific landmarks (subway stations in our case) and their locations on a street map (GoogleMap of London). The efficacy of the participant's learning was evaluated by having them answer a series of questions relating to spatial relationships between the landmarks. Three modes of scrolling were examined: active (A): tilt + button press; marginally active (MA): button press and Passive (P): purely observation. Experiment 1 differed from Experiments 2(a), 2(b) and 3 by the lack of auditory cues. Experiments 2(a) and 2(b) differed from Experiment 3 by the type of sound used (continuous vs. non-continuous).

In Experiment 1, we compared learning performance in the three different modes. In Experiments 2(a) and 2(b), we enhanced the three modes by adding continuous auditory cues that gave information on directions. In Experiment 3, we changed the continuous auditory cues to noncontinuous ones, which gave information on distances and directions.

In the three experiments, we define the landmarks relative to each other. Indeed, two types of distance judgments are classically distinguished: either one asks the participant to locate an absolute directional information (i.e., cardinal directions) or she/he is asked to identify relative directional information (which corresponds to the distance between two objects). It is important to note that in all the modes, all the distances and positions are always allocentric [Klatzky 1998; Farrell et al. 2003].

2. EXPERIMENT 1: ENGAGING ATTENTION FOR MAP EXPLORATION BY USING MOTOR ACTION

The goal of Experiment 1 was to determine whether tying display changes to user motor actions affects learning of spatial relationships between landmarks. Our working hypothesis was that the increased attention to the display, resulting from preparing associated motor acts, will facilitate the learning of geometric relationships among items in a large abstract space, such as maps or even menu trees.

2.1 Method

2.1.1 Participants. Twenty one participants (14 males and 7 females) aged between 18 and 30 took part in this experiment. They were McGill University students and were paid for their participation. They were unaware of the purpose of the experiment and reported no *a priori* knowledge of the map of London. The institutional ethics committee approved the experimental protocol, and participants gave their informed consent before participating.



Fig. 1. Nokia 5500 Sport model mobile phone.

2.1.2 Apparatus. For this experiment, we used the Nokia 5500 Sportphone, as pictured in Figure 1. This phone was chosen because it contains [Nokia 2008] a built-in 3D accelerometer, which provided a means of permitting user actions to control the display on the phone in the experiment.

2.1.3 Materials. We presented participants with a large Google map of London (1254×1292 pixels) on the small display of the phone. Naturally, at any given time, the participants could only see a small portion of the map (208×208 pixels), as depicted in Figure 2. The display would scroll in one of three different modes to 16 different underground stations, one at a time in a random sequence determined by the computer. The participant was instructed to learn as many of the relative locations of these underground stations as possible in a given time frame. After the learning phase, the participant was asked a series of questions intended to gauge how successfully he or she had learned the positions of the stations.

Each participant was tested on one of the following three scrolling modes.

- (1) **Passive (P) mode.** The display was scrolled without any user interaction. The display would simply scroll from station to station, under computer control. Once a station was reached, the scrolling would stop for approximately two seconds, and the phone would vibrate briefly to alert the participant that a station has been reached. The scrolling would then begin to the next station, without any input from the participant.



Fig. 2. Google map of London with phone display serving as a window.

- (2) Marginally active (MA) mode. The display was scrolled automatically under computer control between stations, however, the participant triggered the transition between stations by pressing a button on the cellphone. As in the passive mode, once a station was reached, the phone vibrated briefly to inform the participant that a station has been reached.
- (3) Active (A) mode. The scroll direction is chosen randomly by a computer control, as in the two previous modes, however, the participant initiates scrolling by pressing a button and then tilting the phone in the desired direction, guided by the template, in order to reach the goal. The stronger the tilt, the faster the scrolling. The phone vibrates to indicate that a station has been reached. The participant has to position the phone to the neutral position before initiating the next trial.

In each of the scrolling modes, a peripheral cue provided a visual indication of the direction to the next station. This was intended to direct the allocation of attention. The peripheral cue serves three functions. First, it provides a template of the visual appearance of the target station, which shows the user what he or she is looking for. As with IVD (Interactive Visual Dialog) system [Arbel and Ferrie 2002], where, rather than showing the user how to rotate the object being identified, the user is shown what the desired outcome is. Second, the peripheral cue gives directional cues. The template is placed on the edge or corner of the screen that is closest to the station to which it corresponds. For example, in Figure 3, the template of the “Elephant & Castle” underground station is located on the top left-hand corner of the display of the map. This positioning of the template indicates that the “Elephant & Castle” underground station is located above and to the left of the current display of the map. Finally,

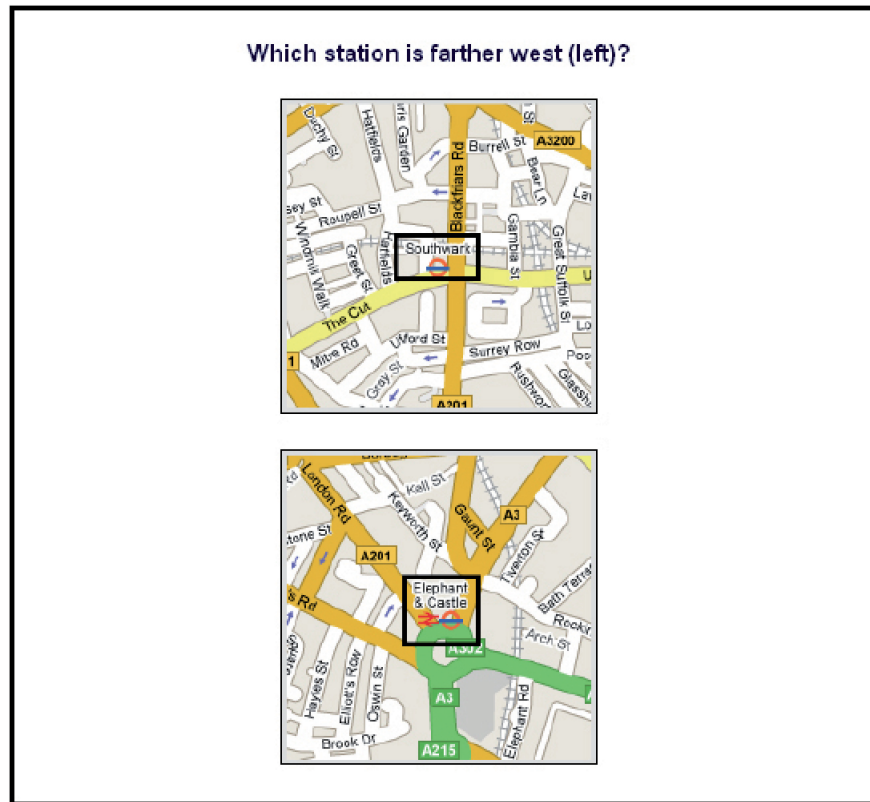


Fig. 4. Question of type 1 for testing participants' learning.

phases six times. Each of the test questions were presented in randomized order such that there were a total of 60 (10 questions × 6 blocks) questions, and each participant completed 300 trials (50 trials/block × 6 blocks).

Before starting the all participants were given time to explore, using the cellphone display, a Google map of the city of Montreal (their home city, and hence they were assumed to be familiar with it) so that they would become accustomed to the controls of the cellphone and with the scrolling modes. This exploration was not a part of the recorded time spent, but only a way to give practice time with the controls. The participants were trained according to their assigned scrolling mode (P, MA, or A) used in the experiment.

Figure 4 shows an example of a question of type 1. In order to avoid biasing the participants toward a particular answer, the two images given in the test were placed deliberately in such a way that their positions with respect to one another would not influence decision making. For example, if the question was attempting to ascertain which of two given stations were farther north (type 1 questions), the images were placed in a row (side-by-side) rather than in a column. In so doing, we avoided the phenomenon where participants were biased to choose the image on top rather than the image on the bottom, since the image on top might appear to be farther north.

2.1.5 *Results.* The mean and standard deviation for participants' answers are presented in Table II. The two-way mixed ANOVA with repeated measures of the "type of question" factor showed

Table II. Descriptive Statistics for Participants' Answers for the Four Experiments

Experiments		Experiment 1		Experiment 2a		Experiment 2b		Experiment 3	
Mode	Type of question	mean (%)	SD	mean (s)	SD	mean (%)	SD	mean (%)	SD
Passive	Left/Right	59.74	0.14	62.52	0.08	61.86	0.17	52.12	0.19
	North/South	65.40	0.17	46.87	0.24	69.86	0.13	58.71	0.09
	Close	54.73	0.22	65.07	0.13	54.29	0.13	56.43	0.21
	Center	52.04	0.19	40.54	0.23	60.14	0.23	50.37	0.17
M active	Left/Right	72.83	0.16	69.84	0.15	80.71	0.08	69.43	0.15
	North/South	80.95	0.11	76.98	0.08	80.71	0.13	76.52	0.16
	Close	73.64	0.16	69.98	0.22	84.57	0.14	74.05	0.16
	Center	68.28	0.15	73.87	0.23	69.86	0.19	81.49	0.12
Active	Left/Right	79.03	0.10	56.98	0.16	62.43	0.18	87.50	0.49
	North/South	84.00	0.13	71.08	0.10	70.00	0.16	88.12	0.71
	Close	71.35	0.13	58.14	0.12	54.57	0.22	89.74	0.38
	Center	90.48	0.12	82.82	0.11	53.14	0.25	87.46	0.11

Table III. Reaction Time (RT) for the Four Experiments

Experiments		Experiment 1		Experiment 2a		Experiment 2b		Experiment 3	
Mode	Type of question	RT (%)	SD	RT (s)	SD	RT (%)	SD	RT (%)	SD
Passive	Left/Right	6.64	1.70	7.11	2.25	6.72	2.86	6.20	1.22
	North/South	5.78	1.90	6.38	1.98	6.26	2.67	6.24	1.76
	Close	8.37	3.61	8.49	3.69	5.23	1.87	6.64	2.23
	Center	7.04	3.96	5.70	1.81	4.87	1.56	5.59	1.47
M active	Left/Right	6.45	0.98	6.06	1.56	9.00	3.11	4.78	0.43
	North/South	6.11	1.82	6.11	2.05	6.80	1.74	4.72	0.80
	Close	7.16	1.51	7.97	2.77	5.32	1.61	6.30	1.67
	Center	5.95	1.21	7.06	3.44	4.74	1.84	4.52	0.77
Active	Left/Right	5.97	1.45	6.25	1.99	6.44	3.48	4.54	1.03
	North/South	5.69	0.93	5.55	1.67	7.79	2.78	4.40	0.87
	Close	8.25	2.61	7.71	2.33	6.69	1.63	5.06	1.42
	Center	5.74	1.30	6.06	1.93	6.30	1.51	4.26	1.08

a significant main effect for the mode factor, $F(2, 18) = 17.20, p < .001, \eta_p^2 = .66$, but a nonsignificant main effect for the question type factor, $F(3, 54) = 1.59, p > .05, \eta_p^2 = .08$ on the participants' answers. The interaction effect between mode and question type was nonsignificant as well, $F(6, 54) = 1.09, p > .05, \eta_p^2 = .11$. *Post hoc* analyses were performed on the mode factor using Tukey's tests to identify exactly where significant differences exist. The analyses revealed that the participants trained in either the active or marginally active scrolling modes showed statistically significant better performance than those trained in the strictly passive mode ($p < .05$). No other differences were found.

Table III shows the mean and standard deviation for reaction time (RT). Only the main effect for the question type factor was significant, $F(3, 54) = 11.15, p < .001, \eta_p^2 = .38$. Pairwise comparisons tests revealed that the question of type III (which is closer to this station) was significantly different from all the other types of questions.

2.1.6 Discussion. The results of Experiment 1 suggest that having an active component does indeed facilitate processing of visual input, as compared with purely passive viewing. Therefore, active exploration of a map on a cellphone may reinforce the spatial knowledge of the participant, as suggested by several previous studies with similar findings [Peruch et al. 1995; Carassa et al. 2002; Christou and Bülthoff 1999]. Peruch et al. [1995] proposed that recall of spatial layout was better in an active compared to a passive condition, as the participants were able to respect perception-action coupling and to

correlate motor and visual inputs. The better performance with active exploration could therefore be due to the optimized perception-action integration particularly required to perform a spatial task.

The reaction time results suggest that participants are faster in detecting absolute positions than relative positions. Indeed, RT for directional questions (left, right, south and north) are shorter than for questions about distance to the closest question. The cognitive process and the neuronal mechanisms involved in relative spatial representations imply that a transformation is applied to absolute spatial representations [Crowe et al. 2008] and then increase the reaction time for relative positions.

3. EXPERIMENTS 2(A) AND 2(B): EFFECT OF CONTINUOUS AUDITORY CUES ON VISUAL-SPATIAL LEARNING AND MOTOR ACTION

Sounds are often used to give environmental information as they travel and are detected by the auditory system of people moving about in the world. As we are surrounded by sound in our daily lives, at every moment, a continuous sound feedback intuitively seems to be an appropriate means of providing an efficient interaction with the device and in assisting the user in finding his way on a map. In experiments 2(a) and 2(b), we tested whether, by adding continuous auditory cues, performance in the active and the marginally-active map scrolling modes would still be better than in the passive exploration mode, and whether the active mode would differ from the marginally-active mode. The goal was to show the possibility of increasing attention to the display by adding audio cues, in order to learn the landmarks' relationships on a map while the participant is completely active. The same scrolling modes used in Experiment 1 were used and augmented with a continuous audio location cue.

3.1 Method

3.1.1 Participants. For experiment 2(a), 21 participants (9 males and 12 females), aged between 18 and 41 (mean = 23, SD = 5.14), took part in. They were McGill University students and were paid for their participation. For experiment 2(b), 21 participants (9 males and 12 females), aged between 17 and 41 (mean = 21, SD = 7.15), took part. They were Wilfrid Laurier University students and were paid for their participation. For both experiments, the participants were unaware of the purpose of the experiment, and they reported no *a priori* knowledge of the map of London.

3.1.2 Apparatus and Materials. The mobile phone used was the same as that in Experiment 1. The map used in this experiment was also the same, and the motor actions for the three modes were the same as in Experiment 1. The only difference with Experiment 1 is that, in this experiment, an audio signal was played continuously during the exploration task. The audio signal varied as a function of the absolute location of the screen center on the global map, and thus gave absolute information about the current location. The audio signal was generated in the software on the cellphone using a Nokia smartphone Symbian OS API routine, called CToneAdapter. This routine takes the frequency and duration of the desired sound as input and generates a continuous series of short tone bursts. The duration of the tone bursts was 70 milliseconds.

The global map was divided into 130 segments (10 rows and 13 columns). The audio signal played in each segment was distinguished from the others by two characteristics: tone burst frequency and tempo (rate of the tone burst repetitions). Using the coordinate system depicted in Figure 5, the tempo of the signal increased along the x-axis, and the frequency of the signal increased along the y-axis for Experiment 2(a). Conversely, for Experiment 2(b), the frequency decreased along the y-axis. The frequency of the signal starts from 500Hz and increases by 100Hz in each segment along the y-axis, resulting in a 1400Hz signal in the last segment. For Experiment 2(a), the tempo of the signal starts from 136 BPM (beats per minute) (moderately quick) and increases by 25 BPM in each segment along the x-axis, resulting in a 461 BPM (extremely fast) signal in the last segment. For Experiment 2(b),



Fig. 5. The tempo of the auditory cue increases on the x-axis from left to right, while its pitch increases on the y-axis from top to bottom for Experiment 2(a) and decreases from top to bottom for experiment 2(b). For example, for Experiment 2(a), the audio signal for the station “Vauxhall” has a high pitch and a slow tempo because it is on the lower left of the map, while the audio signal has a low pitch and a fast tempo for the station “Tower Hill,” which is on the upper right part of the map (black circles). Conversely, for Experiment 2(b), the audio signal for the “Vauxhall” station has a low pitch and a slow tempo, while the audio signal for “Tower Hill” station has a high pitch and a fast tempo.

the tempo of the signal starts from 100 BPM (moderate) and increases by 6.15 BPM in each segment along the x-axis, resulting in a 180 BPM (very fast) signal in the last segment.

3.1.3 Procedure. Experiments 2(a) and 2(b) were run on 21 participants who were divided into 3 groups. As in Experiment 1, each group was trained on a different scroll mode (with each participant executing around 300 trials). After the training, learning of the spatial landmark relationships on the map was tested by asking questions similar to those in Experiment 1. The questions concern the estimation of the spatial position of the stations on the map and their distance relative to the center and/or other station.

3.1.4 Results

—Experiment 2(a): Descriptive statistics for the participants’ responses are shown in Table II. The two-way mixed ANOVA, with alpha of .05, indicated that participants’ answers were influenced by mode (P, MA, A), $F(2, 18) = 5.84, p = .01, \eta_p^2 = .39$, but not by question type, $F(3, 54) = .12, p > .05, \eta_p^2 = .007$. Games–Howell *post hoc* tests revealed that the passive mode differed significantly from active and marginally active modes.

Table III shows descriptive statistics for reaction time (RT). The question type factor was significant, $F(3, 54) = 13.22, p < .001, \eta_p^2 = .42$. The pairwise comparison test revealed that question III (what is closer to this station) was significantly different from all other types of questions. No other significant difference was found.

Finally, one-way ANOVA for independent measures showed that there is no significant difference for passive and marginally active modes for all the types of questions ($p > .05$). For the active mode, the ANOVA showed a significant difference between Experiments 1 and 2(a), for left/right [$F(1, 13) = 10.01, p < .05, \eta^2 = .45$], north/south [$F(1, 13) = 4.54, p < .05, \eta^2 = .27$], and close [$F(1, 13) = 3.91, p < .05, \eta^2 = .24$] type of questions, but not for the center question, $p = .25$. No significant differences were found for RT.

—Experiment 2(b): Descriptive statistics for the participants' responses are shown in Table II. The two-way mixed ANOVA, with alpha of .05, indicated that the participants' answers were influenced by mode (P, MA, A), $F(2, 18) = 6.88, p = .006, \eta_p^2 = .43$, but not by the question type, $F(3, 54) = 2.94, p > .05, \eta_p^2 = .14$. Games–Howell *post hoc* tests revealed that the marginally active mode differed significantly from passive and active modes.

Table III shows descriptive statistics for reaction time (RT). The type of question factor was significant, $F(3, 54) = 12.98, p < .001, \eta_p^2 = .42$. Pairwise comparisons tests revealed that questions I (farther left/right) and II (farther north/south) were significantly different from questions III (closest station) and IV (closest to the center).

Finally, one-way ANOVA for independent measures between Experiments 1 and 2(b) and Experiment 2(a) and 2(b) show that there is no significant difference for passive and marginally active modes for all the types of questions ($p > .05$). For the active mode, the ANOVA showed a significant difference between Experiments 1 and 2(b), for left/right [$F(1, 13) = 5.47, p < .05, \eta^2 = .31$], north/south [$F(1, 13) = 4.84, p < .05, \eta^2 = .029$], close [$F(1, 13) = 5.27, p < .05, \eta^2 = .31$] and center [$F(1, 13) = 12.86, p < .05, \eta^2 = .52$] type of questions. The ANOVA also showed a significant difference for the active mode between Experiments 2(a) and 2(b) [$F(1, 13) = 8.46, p < .05, \eta^2 = .41$] for center types of question, but not for the other types of questions. No significant differences were found for passive, marginally active modes, and RT.

3.1.5 Discussion. For Experiment 2(a), as in Experiment 1, participants who were active and marginally active had better performances than participants who were passive observers, but contrary to our expectations, no significant difference was found in the spatial orientation ability between participants who were active in their exploration of the map of London and participants who were marginally active. The results of the Experiment 2(a) showed that there is no difference between the two modes when continuous audio cues were added. Conversely, in Experiment 2(b), the participants who were using the marginally active mode showed better performances than the two other modes. Similarly to Experiment 1, the participants answered more quickly when the questions were related to the relative position. Husain et al. [2002] showed that a fast tempo (160 bpm) increases the spatial abilities while performing a spatial task. Indeed, participants who used the marginally active mode showed significantly performances when the tempo varied between 100 and 180 bpm (Experiment 2(b)), but not when the tempo varied between 136 and 461 bpm (Experiment 2(a)). Concerning the active mode, the performances dropped significantly when adding a continuous auditory cue. In fact, performance was exacerbated by the additional cues when the participants were active in Experiments 2(a) and 2(b). We believe that participants' attention was split by the tempo BPM while performing the tilt action, and this may have distracted some participants and caused them to lose focus while they were tilting the phone. Besides, the continuous auditory cue for moderate to fast tempos proved to be effective when the tilt was not necessary and when the action consisted of a simple key press.

Table IV. Specifications of Sound Signals for the Noncontinuous Auditory Signal

Signal No.	Frequency (Hz)	Duration (msec)	Distance (pixels)
1	1400	10	<250
2	1100	340	<500
3	800	670	<750
4	500	1000	>750

4. EXPERIMENT 3: EFFECT OF NONCONTINUOUS AUDITORY CUES ON VISUAL-SPATIAL LEARNING AND MOTOR ACTION

In as much as no performance differences were observed in Experiment 2(a) between the MA and A modes, it is important to examine what particular aspects of exploration lead to an efficient multimodal interaction without any overloading of a modality on the other. In the third of our experiments, instead of using an audio signal that is played continuously during the experiment (as in Experiments 2(a) and 2(b)), we used a transient noncontinuous sound (beep) which gives relative information on the station position. This sound is played only during a transition from one station to the other, that is, during the tilt. In this manner, the sound is associated directly to the participant's action. The frequency variation and the sound duration are determined by the distance between two stations: the further apart the stations, the higher the pitch and the longer the duration of the sound.

4.1 Method

4.1.1 Participants. Twenty one students from McGill University (8 males and 13 females), aged between 19 and 35 (mean = 24, SD = 5.77), participated in this study and were compensated. They were unaware of the purpose of the experiment and they had no a priori knowledge of the map of London.

4.1.2 Apparatus and Materials. The equipment and map were the same as those used in Experiments 1 and 2. This experiment differed by adding transient noncontinuous auditory cues that provided information about distance between the current fixation location and the target fixation location. In this scheme, first the distance between the target location and the current location was calculated and then the audio cue corresponding to this distance was played before moving to the next location. Four different audio signals were used in this experiment. The specifications of these sounds are displayed in Table IV.

In this table, the frequency and duration of different signals as well as the radial distance (in pixels) for the corresponding cue is shown. As mentioned, before the distance was calculated as the Euclidian distance between the current location and the target location coordinates (the whole map of London is 1254×1292 pixels). The signs before the distances in the table show the effective range of that signal. For instance, signal 1 was played if the distance between the target and the current location was less than 250 pixels, and signal 4 was played if this distance was more than 750.

4.1.3 Procedure. The experiment was run on 21 participants who were divided into three groups. As in Experiments 1 and 2, each group was trained on a different scroll mode (each participant performed about 300 trials). After the training, the learning of the spatial relationships between landmarks was tested by asking the same types of questions as those in Experiments 1 and 2.

4.1.4 Results. Table II displays the descriptive statistics for participants' correct answers. The two-way mixed ANOVA with repeated measures of the question type showed a significant effect of the factor mode, $F(2, 18) = 52.32, p < .001, \eta_p^2 = .85$, but no significant effect of the factor "type of question,"

$F(3, 54) = .44, p > .05, \eta_p^2 = .02$. Games–Howell’s *post-hoc tests* showed there were significant differences between the performance of the participants who completed the active condition, the performance of those who completed the marginally active condition, and the performance of those who completed the passive condition ($p < .05$). Indeed, the results, shown in Table II, show that the participants trained in the active and marginally-active scrolling modes showed better performance on all four types of questions than the participants trained in the strictly passive mode. Moreover, the participants who trained in the active scrolling mode showed higher performance levels than the participants trained in the marginally active mode for all types of questions.

Descriptive statistics for RT are displayed on Table III. The ANOVA analysis showed no significant difference for the factor “mode,” $F(2, 18) = 1.42, p > .05, \eta_p^2 = .28$, but a significant difference for the factor “type of question,” $F(2.01, 36.16) = 9.68, p < .001, \eta_p^2 = .35$. Pairwise comparisons showed that the question type III was significantly different from all other questions, ($p < .05$).

Finally, one-way ANOVA for independent measures showed that the difference in the results of Experiments 1 and 3 is significant for the active mode for left/right [$F(1, 13) = 4.31, p < .05, \eta^2 = .27$], north/south [$F(1, 13) = 4.62, p < .05, \eta^2 = .45$], and close [$F(1, 13) = 12.94, p < .05, \eta^2 = .52$] type of questions, but not for the question, $p = .64$. It also showed that the results of Experiments 2(a) and 3 are significantly different for the active mode, for left/right [$F(1, 13) = 23.90, p < .05, \eta^2 = .66$], north/south [$F(1, 13) = 13.92, p < .05, \eta^2 = .54$], and close [$F(1, 13) = 43.83, p < .05, \eta^2 = .79$] type of questions, but not for the center question, $p = .45$. Finally, the results showed significant differences between Experiments 2(b) and 3 for the active mode, for left/right [$F(1, 13) = 15.62, p < .05, \eta^2 = .56$], north/south [$F(1, 13) = 13.49, p < .05, \eta^2 = .53$], close [$F(1, 13) = 39, p < .05, \eta^2 = .76$], and center [$F(1, 13) = 11.30, p < .05, \eta^2 = .48$]. No significant differences were found for RT. Except for positioning the center, the noncontinuous sound gave very effective information about the absolute and relative positioning of the stations for the active mode.

4.1.5 Discussion. Experiment 3 demonstrated that active and marginally active exploration with a noncontinuous sound leads to improved learning of spatial relationships. We hypothesize that this is due to an increased level of attention the display caused by coupling the action and the audio cue to the display. When people navigate actively, they encode the visual information of the spatial layout, the direction of the tilt, and the auditory information which are available simultaneously. However, the discrete sound had no benefit for center detection. Since all the tilting movements toward the center can be started from the four cardinal directions (i.e., all the stations in the center can be reached from the four cardinal directions), the sound does not have a significant effect on performance. The tilting seems to be sufficient to locate the center of the map. Concerning the other directions, the availability of relevant and efficient multisensory information may enable participants to develop a stronger and more accessible spatial knowledge and allow integration of this available spatial information. In contrast, the absence of proprioceptive cues and a nonconcordant sound (as in A and MA, where participants had to press button to generate the sound) may oblige the participant to encode the spatial information discretely and relatively to the events he or she received. Conversely to Experiments 1 and 2(a), where no significance difference was observed between the active and the marginally active modes, an active exploration led to a significantly improved performance compared to the marginally active exploration. Adding auditory cues that interact efficiently with the active cues may increase visuospatial learning and help maintain the participant’s attention focused on the device. Finally, allocentric information is integrated faster in Experiment 3 than in Experiment 1. As the sound is associated to relative radial distances, this might help participants to associate the relative positions of the stations to the corresponding frequency.

5. GENERAL DISCUSSION

We have studied the possible enhancement of visuospatial learning obtained by the addition of auditory cues and the coupling of visual changes to motor exploration. The present study is the first to compare active learning to passive learning in the context of spatial orientation on a mobile phone. It was demonstrated that participants developed a deeper spatial knowledge when learning their environment actively than passively. Our experiments suggest that, in active exploration modes, the learning of geometric relationships among landmarks on a large space is facilitated by the participants' attention to the display. Essentially, the audio signal was used to provide the user with information about the location of the current screen on the map. The information that the audio signal gave to the user was divided into relative and absolute information about the current screen location on the map. In Experiments 2(a) and 2(b), the audio signal was played continuously during the exploration, while in Experiment 3 it was played only during a transition from one station to the other. The two experiments suggest that the mere addition of auditory cues does not necessarily enhance the visuospatial learning of the map, especially if the sound forces the participants to concentrate on the task and/or disturb them during the task. However this sound could be efficient for a simple action task (i.e., a key press) during a spatial learning. The experiments also suggest that efficient auditory cues, which do not distract from the other senses, improve performance for an active exploration and create an efficient means of multimodal interaction. Indeed, Experiment 1 suggests that while interacting with a mobile device to obtain orientation on a map, active exploration helps participants during their spatial learning, as attention is focused on the directions of the tilt and then towards the stations displayed in these directions. Experiment 3 confirms these results, as the sound cue gave punctual information about the relative directions as well as information about the absolute position of the stations and/or the center. This short and efficient sound helped the participants to integrate the information at the moment of the action (pushing a button for A and MA); the absence of an active command in the passive condition did not allow this integration. In Experiments 2(a) and 2(b), the degradation of performance in the active mode can be explained by the incompatibility of the continuous auditory cue with the tilting. Indeed, the relation between the two might have created a cognitive overload because the user was not able to perceive and interpret all the entities that are relevant for performing the task. The performances of the MA mode for moderate to fast tempos corroborate this conclusion, since participants were able to efficiently exploit the multimodal interaction while performing the task. However this effect disappeared for the extremely fast tempo. Husain et al. [2002] showed that spatial abilities improve with fast tempos as compared with slow tempos (60 BPM). These results suggest that performance of the spatial task is superior with a fast tempo rather than a slow tempo; but they also suggest that there is a maximum limit threshold of the tempo, which then gives the reverse result, and instead of helping the participants to perform the task better, it annoys and disturbs them during the task.

In summary, the results of the three experiments in this article point to the importance of active exploration for acquiring spatial knowledge on small display screens. Modifying and adding efficient auditory cues can improve attention to the display and increase the efficacy of multimodal interaction. These results also suggest that under some conditions (i.e., continuous sound with an extremely fast tempo, continuous sound with tilting), the spatial abilities of active exploration continue deteriorating, while continuous sound with moderate to fast tempo helps the acquisition of spatial knowledge for simple actions such as a key or button press.

REFERENCES

- ARBEL, T. AND FERRIE, F. P. 2002. Interactive visual dialog. *Image Vision Comput.* 20, 9-10, 639–646.
- BREWSTER, S. 1998. Using nonspeech sounds to provide navigation cues. *ACM Trans. Comput. Hum. Interact.* 5, 3, 224–259.
- ACM Transactions on Applied Perception, Vol. 9, No. 1, Article 5, Publication date: March 2012.

- CHATTING, D. J. 2008. Action and reaction for physical map interfaces. In *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction (TEI'08)*. 187–190.
- CARASSA, A., GEMINIANI, G., MORGANTI, F., AND VAROTTO, D. 2002. Route and survey descriptions of paths: The effect of experience of a large-scale environment. *Bull. People-Environ. Stud.* 20, (Special Issue on Spatial Cognition.) 15–17.
- CHRISTOU, C. G. AND BULTHOFF, H. H. 1999. View dependence in scene recognition after active learning. *Memory Cognition* 27, 6, 996–1007.
- CROWE, D. A., AVERBECK, B. B., AND CHAFEE, M. V. 2008. Neural ensemble decoding reveals a correlate of viewer- to object-centered spatial transformation in monkey parietal cortex. *J. Neuroscience* 28, 5218–5228.
- DARKEN, R. AND SIBERT, J. 1993. A toolset for navigation in virtual environments. In *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, 157–165.
- DODIYA, J. AND ALEXANDROV, V. N. 2008. Use of auditory cues for wayfinding assistance in virtual environment: Music aids route decision. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*. ACM, New York.
- DONG, L., WATTERS, C., AND DUFFY, J. 2005. Comparing two one-handed access methods on a PDA. In *Proceedings of the Mobile-HCI'05*. 235–238.
- DUCHAMP, D., FEINER, S. K., AND MAGUIRE, G. Q. 1991. Software technology for wireless mobile computing. *IEEE Network Mag.* 5, 6, 12–18.
- FARRELL, M. J., ARNOLD, P., PETTIFER, S., ADAMS, J., GRAHAM, T., AND MACMANAMON, M. 2003. Transfer of route learning from virtual to real environments. *J. Exper. Psychol. Appl.* 9, 4, 219–227.
- FORMAN, G. H. AND ZAHORJAN, J. 1994. The challenges of mobile computing. *IEEE Computer* 27, 4, 38–47.
- GUIARD, Y., BEAUDOUIN-LAFON, M., AND MOTTET, D. 1999. Navigation as multiscale pointing: extending Fitts model to very high precision tasks. In *Proceedings of the ACM Conference on Human-Computer Interaction (CHI'99)*. ACM, New York, 450–457.
- HINCKLEY, K., PIERCE, J., SINCLAIR, M., AND HORVITZ, E. 2000. Sensing techniques for mobile interaction. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM, New York, 91–100.
- HUSAIN, G., THOMPSON, N. W. F., AND SCHELLENBERG, E. G. 2002. Effects of musical tempo and mode on arousal, mood, and spatial abilities. *Music Percept.* 20, 2, 151–171.
- ISHIKAWA, T., FUJIWARA, H., IMAI, O., AND OKABE, A. 2008. Wayfinding with a GPS-based mobile navigation system: A comparison with maps and direct experience. *J. Environ. Psychol.* 28, 74–82.
- KLATZKY, R. L. 1998. Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In *Spatial Cognition, an Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, 1–18.
- KRUGER, A., ASLAN, I., AND ZIMMER, H. D. 2004. The effects of mobile pedestrian navigation systems on the concurrent acquisition of route and survey knowledge. In *Proceedings of the Mobile HCI*. 446–450.
- LOKKI, T., GROHN, M., SAVIOJA, L., AND TAKALA, T. 2000. A case study of auditory navigation in virtual acoustic environments. In *Proceedings of the 6th International Conference on Auditory Display*. International Community of Auditory Display, 145–150.
- MUNZER, S., ZIMMER, H. D., SCHWALM, M., BAUS, J., AND ASLAN, I. 2000. Computer-assisted navigation and the acquisition of route and survey knowledge. *J. Environ. Psychol.* 26, 300–308.
- OAKLEY, I., ANGESLEVA, J., HUGHES, S., AND O'MODHRAIN, S. 2000. Tilt and feel: scrolling with vibrotactile display. In *Proceedings of the Eurohaptics'04*. 316–323.
- NOKIA 2008. Nokia UK - nokia 5500 - Technical specifications. <http://www.nokia.co.uk/A4221004>.
- PERUCH, P., VERCHER, J. L., AND GAUTHIER, G. M. 1995. Acquisition of spatial knowledge through visual exploration of simulated environments. *Ecological Psychol.* 7, 1, 1–20.
- REKIMOTO, J. 1996. Tilting operations for small screen interfaces. In *Proceedings of the ACM Symposium on User Interface Software and Technology*. ACM, New York, 167–168.
- TRAN, T. V., LETOWSKI, T., AND ABOUCHACRA, K. S. 2000. Evaluation of acoustic beacon characteristics for navigation tasks. *Ergonomics* 43, 6, 807–827.
- WALKER, B. AND LINDSAY, J. 2003. Effect of beacon sounds on navigation performance in a virtual reality environment. In *Proceedings of the 9th International Conference on Auditory Display*. Boston University Publications, 204–207.
- WILLIS, K., HOELSCHER, C., WILBERTZ, G., AND LI, C. 2008. Spatial knowledge acquisition with mobile spatial information: implications for navigation applications. *Computers Environ. Urban Syst.*, Elsevier Press.

Received August 2010; revised April 2011; accepted April 2011