In-Air Spatial Shortcuts : Leveraging Spatial Memory for Command Selection

Yoann Bourse
Ecole Normale Superieure
Paris, France
Yoann.Bourse@ENS.fr

Eric Lecolinet
Telecom ParisTech
Paris, France
eric.lecolinet@telecom-paristech.fr

ABSTRACT

The rise of computer-mediated living keeps adding numerous functions to the home media center. Efficient access and memorization of a wide number of commands is therefore required. We leverage spatial memory to provide interactions enabling fast memorization of a big number of items.

We introduce two shortcut management systems designed to enable micro-interaction in a couch-interaction setting. The first one is an adaptation of Marking Menus to in air directional interaction. The second one is a novel interaction technique relying on deictic location pointing in which users assign the functions to objects in their environment, following a personal more or less abstract mapping.

We first developed and analyzed an environment-based pointing system based on Microsoft Kinect depth camera, and then studied the memorization capabilities offered by those interactions. We push the limit of memorized items to 22 on average for only 3 presentation per item.

Author Keywords

Spatial cognition; memorization; shortcut; in-air interaction; couch interaction; deictic pointing; Method of Loci; Microsoft Kinect

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: Input devices and strategies

General Terms

Design; Experimentation; Human Factors; Measurement

INTRODUCTION

Computer systems are spreading and covering more and more aspects of our lives, enriching the home environment of numerous functions. Smart televisions also provide new functions to the home environment [5] such as internet navigation and applications (Amazon, Youtube, sports, magazines, radios...), games, social functions [7]... Indeed, the home media-center is growing far beyond its basic multimedia functions, which are also growing in number and diversity. Meanwhile, home automation keeps developing increasingly fast, and devices in the home are becoming the control hub for computer-mediated living [6], gaining more and more abilities as technology progresses ensuring smarter power consumption and home security (temperature, lights, locks...) [12].

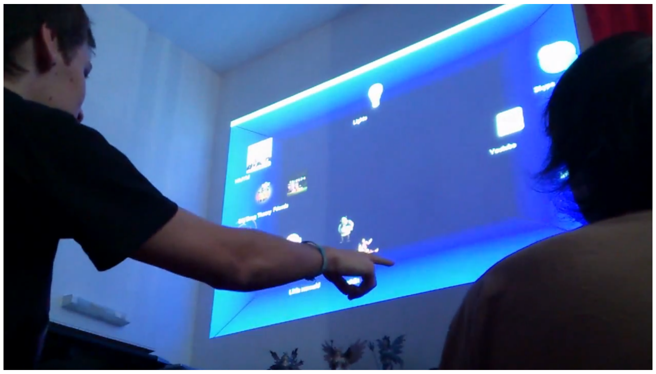

Figure 1. Spatial Pointing Shortcuts associate commands with real-world items

Numerous functions are therefore accessed in a home-environment, for instance from the living-room in a "couch-interaction" setting. It is therefore important to provide users with an easy access to as many of them as possible. Moreover, several of those functions can be repeated a lot [25]. Hence the need for fast shortcuts that can be easily memorized.

To answer this problem, we found inspiration in well-known works in cognitive science [2] and techniques such as the method of loci, which has been used through history, and especially before printing, to learn big number of items with robustness (less recall errors) by leveraging spatial memory [31]. Adapting those techniques to new interactions should allow a big interactional bandwidth at low cost for the user. Spatial memory has also been leveraged in several techniques such as the Marking Menus [19] to ease the learning process, providing a seamless and easy transition between novice and expert mode.

We propose to harness the benefits of spatial memory through two micro-interaction [30] techniques for home environments based on spatial memory and proprioception. Our main contribution is those two techniques, which enable fast memorization of a big number of shortcuts by creating mappings between the shortcuts and either locations or directions in the user's environment.

Since we focus on shortcut managment and not navigation, our interactions are fast, rare and sporadic. This allows us to use in-air interaction without worrying about the physical fatigue which usually plagues a lot of in-air techniques [?]. As an additional challenge, we also tried for our techniques to be usable and pleasant in the context of couch interaction. First of all, the familiarity with the real case environment of our system is expected to boost the performances of spatial memory. Moreover, in-air interaction in such a setting will be convenient: no additional devices are required (like a remote controller which can be lost), allowing for direct interaction and multiple users. In expert mode, no display are required, allowing for interaction with a display turned off (which could be convenient for non-multimedia applications) or without grabbing hold of the display, which may bother people in the room.

As input for our system, we use Microsoft Kinect depth camera [18], as it is already spread-out. The relatively poor precision of our depth camera also raises the additional side question of efficiently inferring the environment from an imprecise and partial input. On a broader perspective, our whole system relies on mapping between different "spaces". The heart of the memorization mechanism is the abstract mapping between the symbolic space inside the user's mind over which we have neither control nor knowledge, towards the space corresponding to their perception of their environment. By their actions, users will create a link between their perceived space and the 3D real space around them. The last mapping happens between this real space and the restricted space that the depth camera perceives. Each of those mappings is in fact a projection between two spaces, coming with a severe loss of information. An additional problem tackled by our project is therefore to infer the original information (in the user's mind) through a very imprecise, noisy and highly fractional projection of it.

In what follows, we will present and analyze the theoretical foundations of our work. We will first introduce and evaluate our second contribution: a pointing system efficient to infer the environment from an imprecise and partial input. We will then describe the micro-interaction techniques that we designed and the experimental study we did to validate their memorization performances.

RELATED WORK

Spatial cognition

Our goal is to leverage spatial memory, which has been known to play a major role in performance in user interfaces. Psychology literature on the benefits of spatial representation for learning is numerous [2]. Spatial learning is known to happen even without focused attention [1], and to strongly correlate with efficiency in computer-system manipulation: Eagan and Gomez [15] are one of the earliest such examples and show that spatial aptitudes are crucial to the manipulation of a software, here a document editor. Numerous studies also highlight the importance of spatial cognition in HCI performances (see [9] for additional references).

We want to exploit the big capacity of spatial memory. Yates describe in The Art of Memory [31] mnemonic techniques, such as the method of loci, used through ancient history. These methods harness the power of spatial cognition in order to memorize a big number of items. They were used in particular before printing to learn important amount of data. For instance, ancient Greeks and Romans memorized law texts by associating each one of them to a precise spatial location, for instance a stone in the layout of a familiar building. By clustering similar concepts and organizing the items, they managed to reach impressive memory capacity [31]. Such methods have been praised for their efficiency and studied by cognitive scientists. We do not aim at memorizing such an important amount of data, but we have good hope that those techniques will provide us with an easy memorization of a still relatively big amount of items.

However, only few interaction techniques attempt to leverage spatial cognition to provide the user with a lot of easily memorizable items. Besides the Marking Menus [19] and the Desktop metaphore [?], the most notable example of this is the Data Mountain, developed by Microsoft Research [26]. The participants had to organize and then retrieve 100 Internet Explorer favorites, using both classical hierarchical menus or a 3D plane on which they put thumbnails of the webpage. Spatial memory allowed for faster selection with less errors and failures in the latter system. Moreover, they also highlighted the durability of spatial memory, as the participants came back 4 months later and showed no significant loss of performance [13].

This work has been carried on by Cockburn and McKenzie [9] and Tavanti and Lind [28] who measured the importance of a 3D spatial representation to help memorization. However, all those studies are limited to a 3D representation on a 2D screen, using a 2D classical mouse interaction, creating a gap between manipulation and representation. To the best of our knowledge, no work attempted to leverage 3D space memory using the new 3D interaction means. It is our guess that the direct correspondence between interaction and representation will enhance the benefits of spatial cognition. Moreover, such an interaction relying on pointing can also benefit from the proprioceptive memory extending the benefits of spatialization [10].

Finally, spatial cognition might help drawing associations between the familiar real world environment and our virtual functions. Associating commands to spatial locations creates powerful mappings through robust memorization : Gustafson et al. [16] have obtained very good performances using this method by allowing users to press imaginary buttons on their hand as if it were their phone's homescreen. In the same way, our system can be seen as an imaginary interface (a shortcut map) superposed to the environment.

Our methods rely on associations between items in the user's environment and the functions of our system. Although those mappings can derive from a semantic link between the item and the function, most of them are arbitrary and abstract, without any straightforward rational explanation (see [29]). Although abstract mappings can seem like unreliable links, it has been shown that they perform very well, almost as good as straightforward semantic mappings [24]. It will be all the more so as the user will get to create those mappings themselves, since agentic and choice-based processing is known to enhance memory [8]. Respecting the particularities of each user's own representations should therefore result in a boost in memorization, as they get to pick and constitute themselves the mappings binding the real world to the shortcuts, instead of learning an arbitrary one which make no sense to them. This also accounts for the need of the system to be customizable, as every user will have personal needs within the tremendous number of functions offered by the system. Thereby, we hope to achieve a fast learning of a big number of commands [26].

Transparent novice-expert transition

Spatial cognition could also help to create an intuitive, fast, transparent beginner to expert transition. Providing a fast interaction for experts is an important key to any human-computer interface, and easing the learning of this expert mode is desirable. The most notable technique emerging from these needs are the Marking Menus [19], circular menus relying on a optional visual feedback for novices. By performing the same gesture for a given command, the transition to expert mode is smooth and transparent. Repetition of these spatial gestures makes learning implicit.

We aim at reproducing such a smart novice to transition expert in our in-air interaction techniques. Although many techniques derived from the Marking Menus exist, few attempts have been made at adapting Marking Menus into in-air interaction. Some studies use additional devices such as a phone or a Wiimote [22]. Bailly et al. [3] obtain a good accuracy despite a relatively long manipulation time, but do not study the memorization process. Our work focuses on the multi-stroke menus proposed by Zhao and Balakrishnan [32] in order to benefit from its accuracy to counterbalance the poor accuracy of the Kinect depth camera.

Our work studies the learning process, offering a memorization analysis similar to Octopocus [4] which improves the memorization of the expert mode in the Marking Menus. However, a major difference of our work is that we leverage already existing spatial knowledge of the user's real-world familiar environment, much like CommandMaps [27] does with the knowledge of the virtual space that is Microsoft Office's ribbons interface.

Deictic pointing

We leverage spatial cognition through the act of pointing, which is a rich field at the border between HCI and cognitive psychology. This problem is indeed not as simple as it seems, in particular in a 3D space. Cockburn et al. [10] compare different pointing techniques: projecting the hand on a virtual 2D plan and use it like a mouse, use the hand as a cursor in the 3D space (slow and inaccurate), or selecting with a laser pointer in the hand. This latter model corresponds to a "What you point at is what you get" paradigm [23] that we want to follow.

However, modeling this natural deictic pointing used in daily life to show things to others is a difficult task. Nickel and Stielfelhagen [21] show that head-hand direction has a better precision to estimate the target pointed intuitively by the user than head orientation, finger direction, forearm orientation or shoulder-hand direction. They also come up with a hybrid HMM model taking those measures as input and outperforming them. But the best estimate of the pointing direction is learned by Gaussian process regression [14]. However, considering the low precision of Kinect and the small performance differences between those methods, head-hand direction is a good enough estimate for the pointing direction in our system.

Most pointing studies use hand-held devices and focus on a user-centered frame of reference. It is the case of Virtual Shelves [20], a project studying the accuracy of pointing in various directions of space. The conclusion is that humans are significantly more precise in the vertical plan right in front of them (zero longitude), and that the targets below horizontal plane (negative latitude) are harder to reach. However, contrary to most studies, we want to focus on an environmental frame of reference, in order to leverage a spatial mapping between the real-world environment and our system. To the best of our knowledge, pointing is indeed rarely considered in a frame of reference that is not user-centered. Therefore, we had to come up with our own pointing system to work in the real-world environment.

POINTING SYSTEM

In order for us to use pointing in an environmental frame of reference, the first part of our work consisted in designing a system which could infer the environment based only on our camera's limited view of the world.

Pointing mechanism

To answer this problem, we propose a paradigm called the **sphere paradigm** to estimate the location of the targeted point in the room-based frame of reference. It relies on approximating the room by a virtual sphere encompassing the real room (4 meter diameter). Being an abstract approximation, this model is expected to have poor accuracy but to be easily transferable. The target point is then estimated by the intersection between the simple $[head, hand)$ pointing direction and the sphere model. It is represented in our system and in this paper by the spherical coordinates (latitude θ, longitude ϕ) of its direction relative to the center of the camera's field of view.

To evaluate this approximation, we compare it to the **room paradigm**, in which we model the room as precisely as possible (in our case, by a cuboid), with a preliminary calibration for instance. This baseline estimates the real targeted point on the wall of the room. Note that this method would be very sensible to any movement of the Kinect and to the calibration process. The simple calibration mechanisms we tried (pointing the same points from two positions) performed poorly. Real implementation would require more sophisticated calibration mechanisms based on continuous movement or sweeping the room with the camera [17]. We used in our

Deviation	Sphere paradigm	Room paradigm
To marker	0.213 (43.3)	0.223 (45.4)
Standard	0.056 (11.2)	0.061 (12.2)

Table 1. Average deviation d (rad) and the corresponding size (cm) on a 2m away wall.

Deviation	Sphere paradigm	Room paradigm
To marker	0.317 (65.6)	0.265 (54.3)
Standard	0.047 (9.45)	0.040 (8.00)

Table 2. Same measures when the user is not centered.

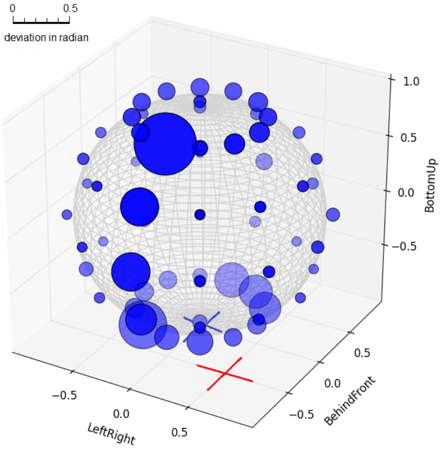

Figure 2. Spatial variation of the total standard deviation using the sphere paradigm with user centered (positioned at the blue cross, contrary to the not-centered condition corresponding to the red cross).

experiment a manually-inputted room model to bypass any calibration bias.

Evaluation

We proceeded to a technical evaluation of our system, in order to estimate its capabilities. In a testing room (6m ; 4.5m ; 2.7m), markers have been placed at 62 points corresponding to all possible latitude and longitude around the center of the camera's field of view considered with a $\frac{\pi}{6}$ step. A user then had to point at these markers and validate the pointing by clicking. We considered two positions for the user: in the middle of the camera's field of view and one big step (85cm) behind on the right. The user uses whichever hand is more convenient and practical so as not to create a biological bias. We measure for each point the spherical coordinates θ (longitude) and ϕ (latitude), and their average deviation θ_d and ϕ_d. We summarize the measures in an average deviation score $d = \sqrt{\theta_d^2 + \phi_d^2}$ for clarity. An average of 30 measures was taken by point.

As we did not use high precision equipment to position the markers, we are fully aware that our manually-positioned markers are very imprecise. Therefore, it comes as no surprise that the deviation to the markers is higher than deviation between different measures for the same point (table 1). This is not a problem as the standard deviation is the most meaningful measure since it can be seen as the deviation from the average point of our measures, that is to say the point which would represent the aimed point inside our system. As a whole, we end up with very satisfactory results, and a precision which would enable any system based on these pointing techniques to discriminate between hundreds of locations.

Problems arise when the hand occlude the head or vice-versa (eclipse phenomenon), in particular behind the user ($\theta = \pi$). We observe a loss of precision for the more extreme values of ϕ: the points are more cluttered, and a small variation of the cartesian coordinates of the body translates into a important variation of angles (fig 2).

Our two paradigm end up having very similar variations. In particular, the sphere paradigm is surprisingly robust to the change of position (table 2). Indeed, when a user points to the same target on the wall from different positions, the same point is measured in the room paradigm, but two different points are measured in the sphere paradigm (the pointing rays intersect on the wall and then diverge). However, it seems that those two points are close enough, because the drop in precision measured when the user is not at the center is relatively similar in the two paradigms. The loss in precision due to the poor accuracy of the camera and the movement of the markers relative to the user outweight the loss of precision due to the sphere abstraction. Moreover, the sphere paradigm provides a robustness and an ease of use missing in the room paradigm. Therefore, by showing similar performances to the precise baseline, our sphere paradigm is smart enough to enable us to fulfill our goal of inferring the environment from the partial and imprecise input data retrieved by Kinect. It will allow us to detect which object is pointed at from any position in the field of view of the camera.

Implementation

The sphere paradigm is the main mechanism allowing us to proceed to real-world pointing from Microsoft Kinect input. Our implementation uses C++ and OpenNI. The code is available upon request. We also designed additional mechanisms to optimize our interpretation of the camera input. Since the depth camera has rather poor precision, smoothing is required to make up for the flickering of the skeleton tracking. We end up with a trade-off between smoothing and precision. Smoothing also causes a delay in the skeleton tracking, which results in the skeleton being always a little behind the actual user silhouette. To circumvent this issue, we added an imperceptible delay (400ms) between the moment where the selection is ordered by the user and the moment it is treated by the system, to give time to the smoothed skeleton tracking to find the right position. We also used a custom Gaussian smoothing algorithm whose effect is exponentially decreasing with the size of the movement. That way, the skeleton tracking followed the user efficiently without delay when he moved from one position to another, but still got rid of the flickering of the detection when precision is required.

INTERACTION TECHNIQUES

We conceived this pointing paradigm to design two in-air micro-interaction techniques: SMM (for Spatial Marking Menu), an in-air adaptation of multi-stroke Marking Menus [32], and a novel interaction technique based on deictic pointing called SPS (for Spatial Pointing Shortcuts). SMM is ex-

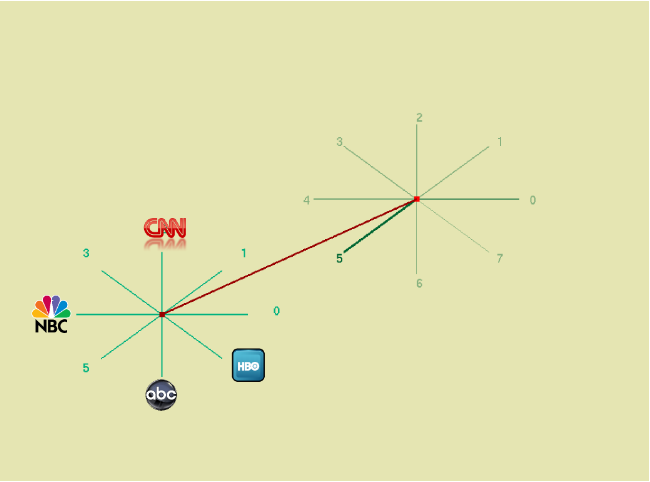

Figure 3. Visual feedback used for SMM (first and second level)

pected to show the well-known good performances of the Marking Menus, and will provide a reference for our novel pointing technique. In essence, SPS relies on pointing a given location (locational), whereas SMM relies on a gesture combining directions (directional).

SMM: Spatial Marking Menus

Spatial Marking Menus (SMM) is an adaptation of the multi-stroke Marking Menus [32] to 3D in-air interaction. It relies on the selection of two directions (two levels of hierarchy) among the 8 canonical ones. Using a two-level Marking Menu allows for a good expressiveness (a total of 64 shortcut storage capacity) and keeping a small interaction. This parameter could however be changed. Each item therefore corresponds to the selection of two branches. It is important to highlight that all those movements can be conceived in the user's frame of reference. They are relative to a starting point of the selection specified by the user by an initial delimiter, allowing for manipulation from anywhere in the room. It also accounts for the diversity of behavior in users.

We compensate for the poor precision of our input by using three clear delimiters to delimit the two segments that will give the directions (see [32]). Although other delimiters could be thought of, such as a snapping of the fingers, we propose to use the brief closing of the hand, fast enough to be efficient, and distinctive enough in order not to create false hits in real-world situations.

Novice mode

The transparent transition from novice to expert mode, like in the Marking Menu paradigm, is enabled by an optional display of a visual help for novices. Help is usually displayed after an inactivity time in the desktop Marking Menus, but it was reported to be annoying and painful to wait without moving during an in-air interaction. We propose to activate the visual help by a voluntary command, such as an audio order ("display SMM") or a gesture from the non-pointing hand.

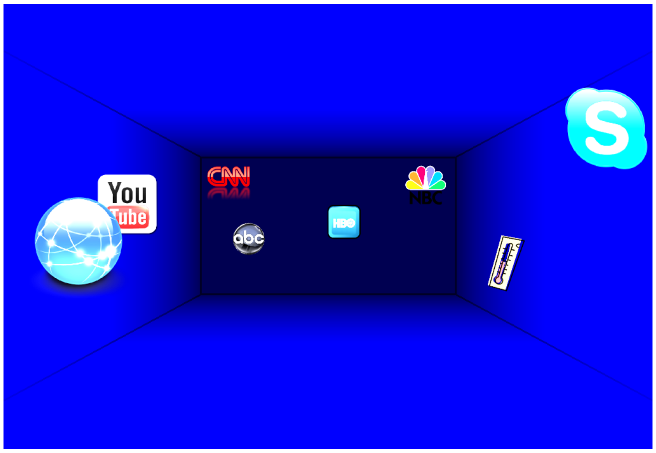

Figure 4. Visual feedback used for SPS

The visual help displayed is similar to the mouse multi-stroke Marking Menus (see figure 3). By construction, it evolves during the manipulation, displaying the first level of menus and then the second level when selected. Exactly like in desktop Marking Menus, novices need to rely on exploration to find what they look for if they don't know where to look. They need therefore to navigate within the menus and cancel their last action, which can be done by a vocal command ("back") or a gesture from the other hand without wasting any interaction possibility.

SPS: Spatial Pointing Shortcuts

Spatial Pointing Shortcuts (SPS) is a novel microinteraction technique allowing a very direct shortcut selection in the context of couch interaction. It relies on direct deictic pointing of the elements of the user's environment, which allows the user to create an abstract mapping between their representation of their real-world environment and the symbolic space of the shortcuts. The selection is straightforward and happens by simply pointing to the desired object and closing the hand without moving. This delimiter seems unnatural enough not to be triggered by accident, especially in a couch setting. If the user points to a place where no item is stored, we decided not to ignore this and to trigger the command corresponding to the closest shortcut.

Novice mode

Much like in SPS, the transition from novice to expert mode relies on the use of on-demand help. SPS relies on a double-level feedback mechanism. An audio feedback indicates the name of the targeted (hoovered) item for precise selection and disambiguation. However, we suggest leaving the possibility for this audio feedback to be deactivated according to the user's preferences.

The visual feedback correspond to the display of a sketch of the room with all the memorized items on it on the nearest monitor. Since our system therefore has no precise information about the environment of the user, we sketch only a rough geometrical representation of the room, displayed as a blue cuboid seen from the inside (see figure 4). This allows us to be compatible with our environment-oblivious pointing mechanism (see *Pointing System*). Our tests show that users understand those limitations and rely on our visual feedback to find the approximate location of shortcuts, or their position relative to each other (which is on the other hand accurate since it does not depend on the environment).

Discussion on our techniques

It seems important to note that our visual feedback does not display the current position of the hand of the user. Continuous feedback seemed indeed to steal the focus of the attention of the user from the performed gesture to the display screen. As our goal was to train users to become experts as efficiently as possible, we wanted them to focus on the actual action to realize in order to learn it better. Thereby, we train the users for an expert mode, which can be used in an eye-free situation, that is to say without even requiring any display. Our tests show that this guess is validated, as users performed poorly in expert mode when continuous feedback was offered.

By leveraging spatial directional perception, SMM provides a well structured environment, hierarchical by design, allowing for easy organization within the shortcuts. This comes at the cost of a complexified manipulation action (several directions choices) and the impossibility to view at once all the recorded items. It is also important to stress out that it is mostly oblivious to the real-world environment.

SPS on the other hand offers a direct access by only a simple action to any shortcut stored in the system. Our precision study (see *Pointing system*) indicates indeed that our system could potentially discriminate bewteen several hundred locations for an important storage capacity. It also provides the user with a highly customizable experience and handles a big variability in the shortcut positioning, allowing the users to have a very personal organization scheme. This makes up for the lack of mandatory hierarchy, as all the items in the system are considered on the same level. The only structure is the one given by the user.

MEMORIZATION EVALUATION

To evaluate our techniques, we designed an within-subjects experiment to measure their memorization capability.

Experimental protocol

Compromises for experimental implementation

For this experiment, we wanted to measure the memorization performance of our interaction technique with as much precision as possible, without being tied to a technological system which could evolve in the future. In particular, in order not to introduce a noise coming from the performance of vocal analysis or closing-hand detection, we used a mouse to simulate all delimiters with a perfect accuracy. The selection delimiters we used were therefore the left clicks. This choice is justified by the fact that we study the memorization capacities of our techniques : coming up with the best delimiter is a whole other problem which would require another study.

We also asked the participants to perform standing to improve skeleton tracking, and allowed them to use whichever hand they felt like using. To emulate a home environment in our laboratory, additional visual cues (pictures of plants, lams, etc...) were added to a testing room to emulate the decor of a living room.

Since we wanted to have a clear measure of the influence of help feedback, we created a clear distinction between novice and expert mode common to the two techniques. Participants would by default enter the expert mode, where manipulation happened without any feedback whatsoever. A right click would trigger the novice mode, that is to say enable all audio and video feedback mechanisms for both techniques once and for all.

To obtain clear measures of memorization and of the performances of our system the experimenter will ask the participant for their intended target in case of selection error. This allowed us to determine if the error comes from memorization or is Kinect-related.

Stimulus

We used in the experiment a neutral vocabulary to represent the commands. It consists of 5 categories (animals, leisure activities, colors, fruits and clothing items). We used 5 items per category, adding up to a total of 25 items. The items were different between techniques, but not between users. They were presented on a big screen. Since we wanted to test the maximal capacity of memory, we decided not to consider the items following a Zipf law but presenting every item the same number of times. Using Zipf law amounts to presenting a few items a big number of times, which overshadows all the other items barely presented at all in an experiment of regular length.

Participants and apparatus

The experiment was taken by a total of 12 participants (3 women), aged from 15 to 30, average 23. Most of them had no previous Kinect experience. The order of the techniques was balanced among participants following a latin square. A two-factor ANOVA on starting phases and techniques with repeated measures on the technique factor (two by subjects) later showed that the order of phases had no significant effect neither on time nor on memory performances, validating this experimental protocol.

We proceeded to the experiment in a test room (6m, 4.5m, 2.7m). The stimuli and visual feedback were displayed on a 40’ screen. The whole experiment lasted about one hour.

Procedure

Each technique test began with a small example phase in order for the participant to get familiar and to understand the manipulation involved. Each user was then asked to chose the location for each item either in the 3D space surrounding them (SPS) or on the two layered directional menu (SMM). Much like in real use, the user therefore gets to pick the location of the items, enhancing memorization (see *Related work*). Moreover, to mimic this real-world use, the participant does not know beforehand what items are going to come in the future, creating a big constraint on their organization scheme.

Retrieval phases then took place, where the participant was asked by a visual and audio stimulus to retrieve the stored items. They were asked to memorize as many items as possible, and to select them as quickly and precisely as possible. An audio feedback lets the user know if they were right or not. The order of the retrievals was randomized every time. Each phase consisted of one retrieval per item. Each technique was tested on 4 retrieval phases, for an overall total of 200 selections by participant. In the first three phases, for every item, the user used by default the expert mode and had the possibility to trigger the novice mode if he judged it necessary. In the fourth phase of each technique, access to the novice mode was disabled, in order to evaluate what had been memorized (after positioning and 3 exposure to stimuli) without any feedback whatsoever.

At the end of the experiment, a survey was given to the participant in order to measure their personal opinion. A small discussion aimed at highlighting the organization strategies and memorization techniques they used, in order to proceed to a user study.

Results

System performances

To better understand the global performances of our techniques as a whole (including the possible novice mode), we distinguish two hit scores as a consequence of our experimental choices. The basic ”hits” score is the number of good selections measured by the system, whereas the ”memory hits” score correspond to the number of correct intended selections, measured by the experimenter as discussed in *Experimental protocol*, that is to say the number of cases where the participant knew where the item he wanted was but did not manage to reach it with the system.

The difference between those two figures corresponds to ”Imprecision errors”, which rely on depth camera manipulation or performances. They can come from a poor skeleton detection from the skeleton tracking (flickering), but also from a bad movement differing from the intent from the user, which are nearly impossible to distinguish. For instance, in SMM, since the human does not manipulate naturally on a vertical plane but on a sphere centered on him whose ray is the length of their arm, it is not rare to see people having trouble to perform a perfectly horizontal movement. It is our guess that smarter detection techniques could be devised.

We compute an average score as the proportion of cases where such imprecision errors happened in the total number of selections. We achieve overall few imprecision errors (11% for SMM, 14% for SPS), with a few outliers (a few people were victims of poor skeleton tracking) dragging the means down.

Memorization

Measure of phase 4 (without any feedback) allows us to assess raw memorization after only 3 exposure by item. From observations, our guess is that incidental learning has also taken place, but there is however no clear way to measure it in our setting. Figure 5 showcases the recall performances of both techniques (and the recall as perceived by the system). We manage to reach with so few exposure events a recall score of 16.4 items for SMM and 22.1 for SPS over a total of 25 items, outperforming desktop memorization techniques. A one-way ANOVA for correlated samples shows that the factor ”technique” has a very significant effect on the memory hits, that is to say the overall number of memorized items ($p = 0.0055$, $F = 11.8$). Therefore, SPS clearly outperforms SMM when it comes to memorization.

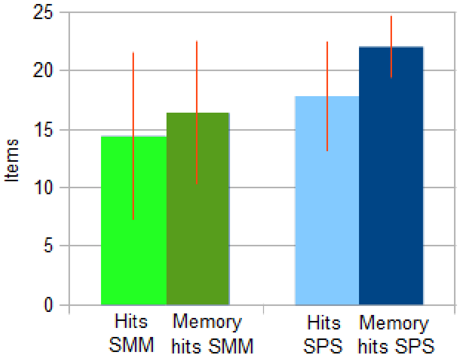

Figure 5. Means and standard deviations of recall performances in expert phase (4)

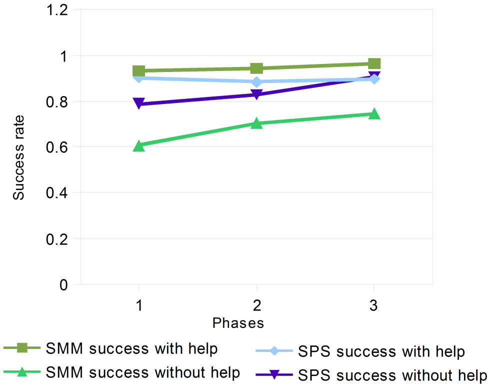

Figure 6. Evolution of success rate on memory hits

Success rates

We study the learning phase in our system through the evolution of the success rate of memory hits (percentage of hits in the total number of selections) over our whole dataset (see figure 6). Performance is quite high from the start, which leaves relatively little room for increase over time.

Errors measured for the ”memory hits” correspond to true errors where the participant did not know where the item was. In expert mode, they were attempts of selection which ended up failed, that is to say cases where the user was mistaken about the item believed location. We also notice some memory errors in novice mode, corresponding to cases where the help feedback was not enough to select the right item. For example, some participants got from the help the approximate location of an item in SPS but failed the selection because

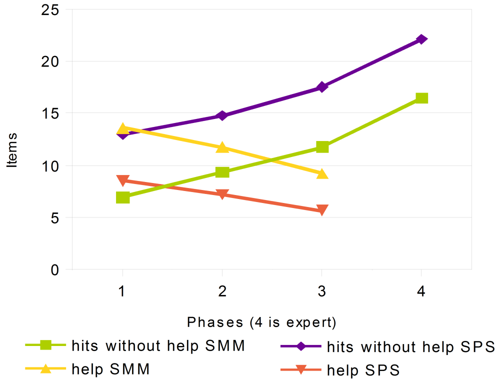

Figure 7. Usage of novice mode and expert hits

they were too much in a hurry to look for the audio feedback. Others ended up selecting a wrong item in SMM by lack of attention. The decrease in those kind of errors shows an increasingly better use of the feedback by the users.

Use of novice mode

The behavior in the use of novice mode (audio and video help) were rather diverse, and two trends seem to appear in our dataset. People unsure of themselves tend to use the novice mode very often, sometimes showing only a very small decrease in help usage. We call this behavior ”timid”. Others are either more risky or feel more comfortable with the techniques right away, resulting in the ”risky” behavior. It is not rare to see people adopting a ”timid” behavior with one technique and a ”risky” one with the other.

Two ANOVA tests (one per technique) showed a very significant effect of our ”help profile” distinction on the novice mode usage ($p < 0.0001$, $F > 60$). Other ANOVA tests showed a significant impact of this profile on memorization ($p = 0.012$, $F = 9.3$ for SMM, $p = 0.011$, $F = 9.8$ for SPS): ’risky’ users are better than ’timid’. Interestingly enough, this ”help profile” did not seem to have a significant impact on the learning rate of the items for SMM ($p = 0.68$) but did for SPS ($p = 0.0038$, $F = 14.1$), implying that the feedback may impact more the learning process for SPS. That being said, the number of participants we had seems too small to draw any significant conclusion.

However, we can still draw conclusions from the averaged behavior. Figure 7 shows the evolution of novice mode usage and good selections in expert mode during the experiment. The two techniques present a smooth and efficient novice to expert transition, which was one of our main concerns. Moreover, this transition is decently fast, for an average of 3.1 items learned by phase, which results in high memorization scores for very few exposure events.

Time

Another measure showcasing the transition from novice to expert is the manipulation time (see figure figure 8). Our system proceeds to two different measures: the total selection time between the apparition of the stimulus and the user’s retrieval, and the reaction time between the apparition of the stimulus and the moment when the system records a significant movement. The space between the two curves corresponds to the time taken by the actual selection movement.

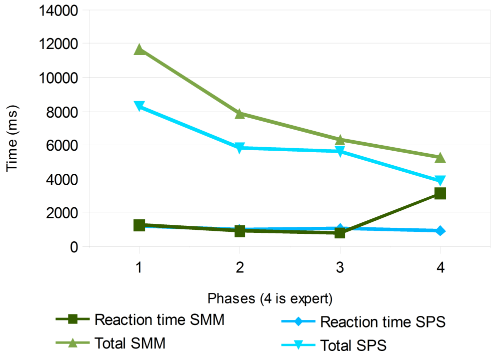

Figure 8. Evolution of reaction and total selection time

	SPS	SMM
I liked the technique	4 *(0.85)*	3.25 *(1.48)*
It was easy to get to grasp with	4.25 *(0.97)*	3.08 *(1.51)*
This technique is fun	4.33 *(0.89)*	3.08 *(1.31)*
This technique is efficient	3.92 *(0.9)*	3.17 *(1.27)*
This technique is fast	4 *(1.13)*	3.75 *(1.22)*
I easily found what I looked for	3.75 *(1.14)*	3.25 *(1.48)*
I learned the items easily	3.75 *(1.14)*	2.83 *(1.34)*
I could memorize a lot of items	4.17 *(0.94)*	3 *(1.41)*
I’m satisfied with my organization	3.58 *(1.56)*	3.42 *(1.38)*
This technique is tiring	2.67 *(1.07)*	3.92 *(1.38)*

Table 3. Means and standard deviation of subjective answers to the user survey (1=strongly disagree, 5=strongly agree). Statistically significant differences are highlighted in bold.

This distinction offers us a very interesting observation. In the expert phase (4), for the SMM condition, a relatively big amount of the time is taken by the ”reaction time”. This highlights a significant hesitation before movement in expert mode using SMM not observed in SPS. A one-way ANOVA on the total time in phase 4 showed that the factor ”technique” has a clear and significant effect ($p = 0.006$, $F = 11.3$). SPS is therefore significantly faster than SMM, ending up at 3888ms against 5267ms for SMM. Note that these times are relatively long, since the experiment was probably not long enough in order for the user to retrieve the items efficiently and fast. However, the fast decrease we observe leaves good hope for improvement over a longer time of use of the technique.

Subjective Preferences and Qualitative Observation

Our experiment allowed us to study the diversity of the personal behavior of the participants in respect to our techniques.

Perception survey

Table 3 sums up the results of the subjective survey distributed to the users at the end of the experiment in order to evaluate their personal perceptions of the techniques. All the values have been tested for statistical significance of the mean differences by a paired difference t-Test for correlated samples, to assess their significance as best as possible considering the small size of our dataset.

Unsurprisingly, user comments highlight that ”the efficiency depends a lot on the initial placement”, and that it is ”hard to organize the items without knowing them beforehand”. In real-world use, the performances of our techniques might therefore be better as the users will have a better idea of what they want to do with them.

SPS has been reported as ”more pleasant”, and ”requiring less reflection”, but also as ”more suited to everyday life”. Those comments bring high hope for our novel interaction technique, in particular in the context of computer-mediated living. It can indeed handle various levels of symbolic abstraction in the mappings involved.

SMM on the other hand was praised for its organizational capacity. It is ”practical to regroup concepts, but not to memorize them”. On the whole, people seem to have liked the constraints of SMM which obligated them to use a decent organization scheme. Comments like ”it is easier to be lost for a big number of items” suggest that this technique might be better suited for small amounts of commands.

Qualitative observation

Observation during the experiences allowed us to notice some recurring trends we weren't able to quantify. There is a great diversity in the positioning scheme of users, even within the different items for the same user. Some participants applied a partial organization scheme which they had to adapt for an unforeseen item.

However, we still can draw interesting conclusions from rough observations. People who apply a clear organization scheme seem to perform better at memorization. Leveraging personal memories or mnemonic devices (inventing stories or semantic proximity) seemed to enhance memorization. This was mostly done in SPS, where people tended to place the items in the environment by a subjective semantic mapping, sometimes oblivious of their category.

For SMM on the other hand, the large majority of participants used the categories suggested by our item taxonomy as first level choice (first direction). For the second level, the default behavior was to place the items in incoming order. As a consequence, they had no trouble finding the right first category, but the second level selection was hard. Some movement combinations seemed easier to remember: twice the same directions (up up), or opposite directions (right left). Some users came up with meaningful organization scheme within categories: semantic proximity, sentiment assignation (down is bad), trying to recreate the shape of the object with the command...

DISCUSSION

Our techniques show efficient memorization, outperforming state of the art of desktop techniques, and show the desired smooth and fast transition from novice to expert mode. SPS is also faster than SMM, and overall preferred by a subjective opinion survey. SPS was therefore very efficient at direct item retrieval, but suffered from poor hierarchical organization. Contrary to the method of loci, few people spontaneously structured their item landscape. SMM showed the exact opposite, being very efficient at providing the user with an organization scheme but performing poorly at distinguishing among the items within a given category. Those two techniques could well be combined to make up for each other's drawbacks.

To that end, we propose as an open perspective to extend them into a more advanced interaction mechanism. This hybrid technique would consist on using SPS in a first level (pointing an object on the environment) and a simple single-level Spatial Marking Menu as a second level (performing a direction selection around this object). This would take advantage of the huge direct retrieval capacity of SPS and the great organization scheme of SMM, in a technique perfectly suited for computer-mediated living. However, this raises other questions (such as the visual feedback to offer in novice mode) and could be the object of further research.

We believe that one of the reasons of our success in memorizing a big number of items, much like the Data Mountain [26], was to leave the choice of the item position to the user (much like in real-case usage), and thereby to enhance their memory through personal implication [8]. However, the ease of participants to create mnemonic techniques from scratch (stories...) with SPS leads us to believe that withdrawing the possibility to choose would mainly handicap SMM.

Finally, many other parameters play an important role on spatial cognition, and by extension on our techniques. Now that they have been introduced, further studies could investigate the influence of the number of visual cues in the environment, of their nature (color, emotional link, etc...), or their organization on the performances of our techniques. They could also study the effect of a larger vocabulary or a longer manipulation time.

CONCLUSION

In order to answer the rise of capability in our home media centers, we leveraged spatial memory to offer a easy memorization of a important number of items. After designing a pointing paradigm inferring the targeted object without knowing the environment nor constraining the user's position, we proposed two in-air micro-interaction techniques. Spatial Marking Menus (SMM) is an adaptation of the multi-stroke marking menus [32], relying on the selection of two directions of space. Spatial Pointing Shortcuts (SPS) is a novel interaction technique relying on deictic pointing to create more or less abstract mappings between the real-world environment and the symbolic space of the shortcuts, following a ”what you point is what you get” intuitive paradigm. Our experiment show that those techniques present a smooth and fast transition from novice to expert mode, and outperform desktop memorization (22.1 items memorized for SPS and 16.4 for SMM) after only 3 exposure to each stimulus. They successful answer our need for easy memorization of a relatively big number of items.

Our system is based on the low-cost depth camera Microsoft Kinect, already present in many households, making our work not only realistic but already applicable. We indeed designed several solutions to improve the performances of this low resolution camera in our context of use. However, the evolution of technology is bound to improve the performances of our techniques with the improvement of depth cameras. Furthermore, the evolution of hardware gives us hope for novel feedback mechanism, with for instance multidirectional projector (or several projectors) or interactive walls. In this context, we could use as visual feedback mechanism a direct projection of the shortcut icon at the real-world position where it is stored (on top of the actual item), eliminating the need for a display. This would bypass the representation issue (3D space displayed on a 2D screen) and may improve the overall performances of the techniques.

On a broader note, we hope to revive interest for spatial memory, and particularly for the very powerful method of loci [31], used through history to learn big number of items. Our work indeed extends the conclusions of the Data Mountain [26] to a 3D input interaction. We showed that the democratization of 3D in-air interaction created new means to leverage existing human powerful capabilities to enhance human-computer interaction, and we hope to lead the way for further research in that direction.

REFERENCES

1. Andrade, J., and Meudell, P. Short report: is spatial information encoded automatically in memory? *The Quarterly Journal of Experimental Psychology 46*, 2 (1993), 365–375.
2. Baddeley, A. *Human memory: Theory and practice*. Psychology Pr, 1997.
3. Bailly, G., Walter, R., Müller, J., Ning, T., and Lecolinet, E. Comparing free hand menu techniques for distant displays using linear, marking and finger-count menus. *Human-Computer Interaction–INTERACT 2011* (2011), 248–262.
4. Bau, O., and Mackay, W. Octopocus: a dynamic guide for learning gesture-based command sets. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, ACM (2008), 37–46.
5. Boxee. http://www.boxee.tv/.
6. Cabrer, M., Redondo, R., Vilas, A., Arias, J., and Duque, J. Controlling the smart home from tv. *Consumer Electronics, IEEE Transactions on 52*, 2 (2006), 421–429.
7. Cesar, P., and Geerts, D. Past, present, and future of social tv: A categorization. In *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, IEEE (2011), 347–351.
8. Cloutier, J., and Neil Macrae, C. The feeling of choosing: Self-involvement and the cognitive status of things past. *Consciousness and cognition 17*, 1 (2008), 125–135.
9. Cockburn, A., and McKenzie, B. Evaluating the effectiveness of spatial memory in 2d and 3d physical and virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, ACM (2002), 203–210.
10. Cockburn, A., Quinn, P., Gutwin, C., Ramos, G., and Looser, J. Air pointing: Design and evaluation of spatial target acquisition with and without visual feedback. *International Journal of Human-Computer Studies* (2011).
11. Company, N. The state of mobile apps, 2011.
12. Control4. http://www.control4.com/.
13. Czerwinski, M., Van Dantzich, M., Robertson, G., and Hoffman, H. The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3d. In *Proc. Interact*, vol. 99 (1999), 163–170.
14. Droeschel, D., Stückler, J., and Behnke, S. Learning to interpret pointing gestures with a time-of-flight camera. In *Proceedings of the 6th international conference on Human-robot interaction*, ACM (2011), 481–488.
15. Egan, D., and Gomez, L. Assaying, isolating, and accommodating individual differences in learning a complex skill. *Individual differences in cognition 2* (1985), 173–217.
16. Gustafson, S., Holz, C., and Baudisch, P. Imaginary phone: learning imaginary interfaces by transferring spatial memory from a familiar device. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, ACM (2011), 283–292.
17. Izadi, S., Newcombe, R., Kim, D., Hilliges, O., Molyneaux, D., Hodges, S., Kohli, P., Shotton, J., Davison, A., and Fitzgibbon, A. Kinectfusion: real-time dynamic 3d surface reconstruction and interaction. In *ACM SIGGRAPH 2011 Talks*, ACM (2011), 23.
18. Kinect, M. http://www.xbox.com/fr-fr/kinect.
19. Kurtenbach, G., and Buxton, W. Issues in combining marking and direct manipulation techniques. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*, ACM (1991), 137–144.
20. Li, F., Dearman, D., and Truong, K. Virtual shelves: interactions with orientation aware devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, ACM (2009), 125–128.
21. Nickel, K., and Stiefelhagen, R. Visual recognition of pointing gestures for human-robot interaction. *Image and Vision Computing 25*, 12 (2007), 1875–1884.
22. Oakley, I., and Park, J. A motion-based marking menu system. In *CHI'07 extended abstracts on Human factors in computing systems*, ACM (2007), 2597–2602.
23. Pan, G., Ren, H., Hua, W., Zheng, Q., and Li, S. Easypointer: what you pointing at is what you get. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, ACM (2011), 499–502.
24. Poage, M., and Poage, E. Is one picture worth one thousand words? *The Arithmetic Teacher 24*, 5 (1977), 408–414.
25. Purcell, K. Half of adult cell phone owners have apps on their phones. *Pew Research Center's Internet & American Life Project. Accessed January 9* (2011), 2012.
26. Robertson, G., Czerwinski, M., Larson, K., Robbins, D., Thiel, D., and Van Dantzich, M. Data mountain: using spatial memory for document management. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*, ACM (1998), 153–162.
27. Scarr, J., Cockburn, A., Gutwin, C., and Bunt, A. Improving command selection with commandmaps. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, ACM (2012), 257–266.
28. Tavanti, M., and Lind, M. 2d vs 3d, implications on spatial memory. In *Symposium on Information Visualization* (2001), 139–145.
29. Wobbrock, J., Morris, M., and Wilson, A. User-defined gestures for surface computing. In *Proceedings of the 27th international conference on Human factors in computing systems*, ACM (2009), 1083–1092.
30. Wolf, K., Naumann, A., Rohs, M., and Müller, J. A taxonomy of microinteractions: Defining microgestures based on ergonomic and scenario-dependent requirements. *Human-Computer Interaction–INTERACT 2011* (2011), 559–575.
31. Yates, F. *The art of memory*, vol. 64. Random House UK, 1992.
32. Zhao, S., and Balakrishnan, R. Simple vs. compound mark hierarchical marking menus. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, ACM (2004), 33–42.