LazyNav: 3D Ground Navigation with Non-Critical Body Parts

Emilie Guy[1] Parinya Punpongsanon[2] Daisuke Iwai[2] Kosuke Sato[2] Tamy Boubekeur[1]
[1]Telecom ParisTech - CNRS - Institut Mines-Telecom [2] Osaka University

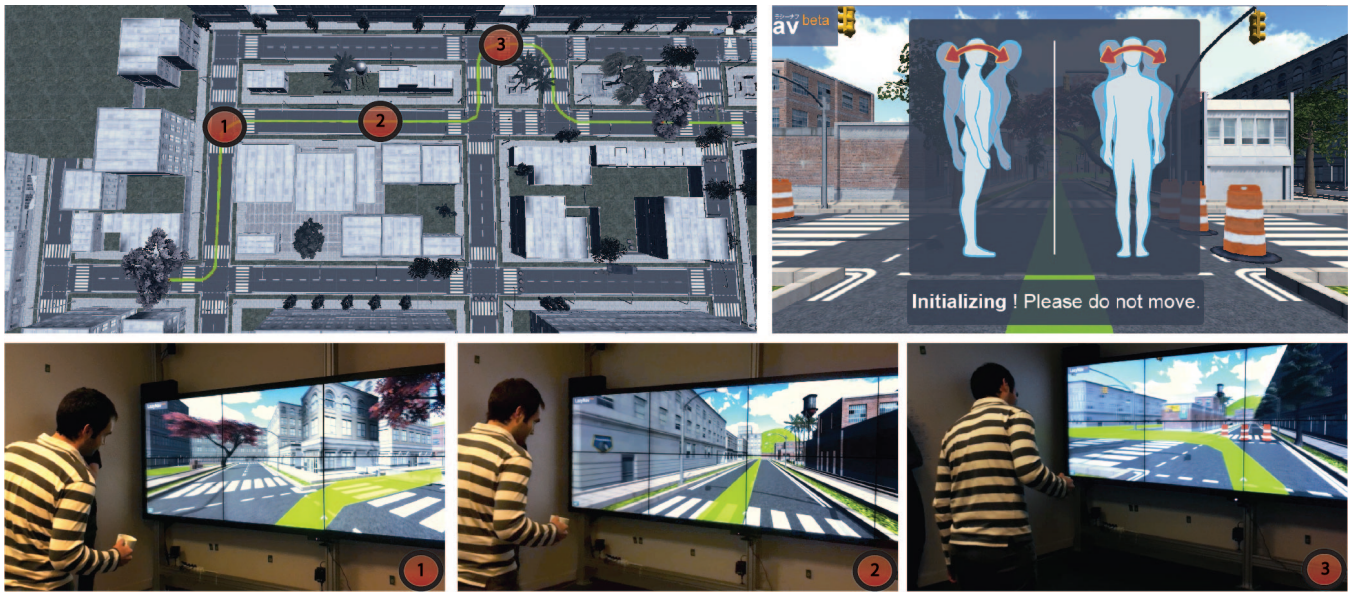

Figure 1: We performed a user study to evaluate several ground navigation metaphors. Top-left: Top view of the virtual city, users have to follow the green path as closely as possible. Top-right: At the beginning of the session, the available motions are displayed to the user, while her rest pose is captured. Bottom: a user is travelling in the scene while holding a cup in his hand.

Abstract

With the growing interest in natural input devices and virtual reality, mid-air ground navigation is becoming a fundamental interaction for a large collection of application scenarios. While classical input devices (e.g., mouse/keyboard, gamepad, touchscreen) have their own ground navigation standards, natural input techniques still lack acknowledged mechanisms for travelling in a 3D scene. In particular, for most applications, navigation is not the primary interaction. Thus, the user should navigate in the scene while still being able to perform other interactions with her hands, and observe the displayed content by moving her eyes and locally rotating her head. Since most ground navigation scenarios require only two degrees of freedom to move forward or backward and rotate the view to the left or to the right, we propose *LazyNav* a mid-air ground navigation control model which lets the users hands, eyes or local head orientation completely free, making use of a single pair of the remaining tracked body elements to tailor the navigation. To this end, we design several navigation body motions and study their desired properties, such as being easy to discover, easy to control, socially acceptable, accurate and not tiring. We also develop several assumptions about motions design for ground navigation and evaluate them. Finally, we highlight general advices on mid-air ground navigation techniques.

Index Terms: H.5.2 [Information Interfaces and Presentation (e.g., HCI)]: User Interfaces—Interaction Styles and Evaluation/Methodology;

IEEE Symposium on 3D User Interfaces 2015
23 - 24 March, Arles, France
978-1-4673-6886-5/15/$31.00 ©2015 IEEE

1 Introduction

Immersive applications benefit directly from the expansion of 3D content, interaction devices and modern displays. With 3D scanners democratization (e.g., Kinect, Leap Motion, or Tango) and online 3D databases, the number of available 3D scenes and applications is growing exponentially. With so much new 3D content, *mid-air interaction techniques* are becoming more and more popular as they provide a new tridimensional way to interact with the virtual content [1], in contrast with classical 2D mapping.

Mid-air techniques provide also a more immersive experience and cope with natural motions. For instance, in the context of public displays, users do not have to connect, touch or wear any specific devices and can instantaneously interact with the system from a distance. Therefore, practitioners envision new public applications where the user navigates inside a 3D scene using simple gestures to visit an historical monument or find her way in a mall for instance.

While 2D devices have their own standards for ground-navigation (e.g., flying and orbiting using the keyboard and the mouse), mid-air techniques still lack natural metaphors for travelling in a scene. More precisely, in many applications, the user must interact with the scene and ground navigation is therefore not the primary action she has to perform. Consequently, the system should keep her hands, eyes or local head orientation completely free and available for interacting with the virtual content (selecting or editing virtual objects) or performing social interactions while playing (i.e. showing content to others).

We observe that ground navigation requires only two degrees of freedom: one for walking (forward or backward) and one for turning the view (rotating to the left or right). Thus we propose

LazyNav, a ground navigation control mechanism which is based on non-critical body parts. We adopt a "lazy" approach e.g., the user easily controls the navigation with non-tiring motions, which are easy to discover, have a fast learning curve and are socially acceptable.

In the following we first explain how we use several body motions that follows the above properties (Section 3). In particular, we make several assumptions mandatory for developing mid-air ground navigation techniques (Section 3.3). Then, we describe the system architecture of *LazyNav* (sec 4) and perform informal and formal user studies to analyze both our motions and our ground navigation assumptions (Section 5). Finally, we highlight general advices for ground navigation design using mid-air interactions (Section 7).

2 Related Work

Navigating in a virtual environment (VR) is a very common scenario and therefore has led to numerous interactions metaphors. The most popular one in desktop configuration is the First Person View (FPV) paradigm found in many video games where the keyboard and the mouse are used concurrently to navigate. As ground navigation using 2D input is outside the scope of this paper (see the work of Jankowski et al. [2] for references), we focus on related works which use 3D user motions to navigate and we classify them in different groups.

Walking-In-Place techniques (WIP): The most straightforward solution for ground navigation might be to have a one-to-one mapping between the user gestures and the virtual motions, i.e. ask the user to actually perform the motions in the real world. However, this solution is restricted by a limited workspace and long or unlimited walks are impossible. Several WIP techniques have been developed to solve this issue. They are usually implemented in immersive environments (e.g., Head Mounted Displays, Cave [3]) and require sophisticated tracking equipments. For instance, *Cyberith Virtualizer* [4] propose a platform where the user can walk, run and jump in place but also turn or squat down. Their goal is to reach as immersive as possible gaming experiences, with user gestures accurately reproduced in the virtual world. Ikeda et al. [5] present an immersive telepresence system to navigate in photorealistic scenes by walking on a treadmill. The aim of this system is to give an immersive sense of walking in a remote site. One-to-one immersive mapping requires non-trivial equipments to ensure the player safety (low-friction base platform, belt system), capture her motions (high precision sensors) and display the virtual world (head-mounted display).

In *Shake-Your-Head* [6], Terziman et al. adopt a different approach: they track the user head movements to control the virtual camera motion. The system is low cost (only a web camera and a standard screen are required), works in a large set of configurations (sitting or standing) and provides different interactions (walking forward, turning, jumping and crowling).

On the contrary to our work, WIP techniques try to mimics as much as possible the user locomotion. This certainly gives a better immersive feeling, however it also results in more tiring and sub-efficient interactions. We rather focus on full-body interactions that are as natural as possible, but also effective and which allow the user to perform a secondary action while navigating.

Desktop Configurations: A number of methods exploit hand gestures to navigate in a 3D scene. Using a leap motion sensor, Adhikarla et al. [7] mimic well-known touchscreen gestures (rotate, pan and zoom), whereas Nabiyouni et al. [8] evaluate several travelling metaphors (air plane and camera-in-hand) and multiple ways to control the speed (discrete and continuous).

Simeone et al. [9] have a different approach, more related to our work. Their key concept is to use only the lower body part of the user for ground navigation. They work in a desktop environment where the user is seated in front of a computer and use her foot to control the navigation. We address different scenarios and target public environment where the user is standing in front of a screen rather than seating at a desk.

General Body Motions: Several techniques use general body motions to navigate in a 3D virtual environment. Daiber et al. ([10], [11]) developed a system mixing multitouch and feet gestures. On the contrary to WIP techniques that preserve the user proprioception, Pettré et al. [12] propose to preserve the user equilibrioception by performing leaning motions to navigate. They use a simple articulated platform on which the user is standing to help her leaning in the desired direction. Although freehand interactions could be achieved by this immersive device, the user cannot perform a secondary action while travelling in the virtual scene.

Ren et al. [13] use a freehand (no hands-on device) gestural technique, with a broom metaphor to travel in a 3D scene: the user hands control the walk and her shoulders control the view rotation. In one of their experiment they give a real physical device (i.e., a broomstick) to help users understanding and performing gestures. In the work of Roupé et al. [14], users lean the bust forward/backward to walk, rotate the shoulders to turn, and raise their arm to stop the motions.

In these approaches, leaning the bust or rotating the shoulders seem to be natural interaction choices, however no common interaction is especially defined for the walk. We believe the set of motions used for ground navigation can be deeper analyzed. Therefore, we evaluate several body motions to understand why some of them are easier to perform, understand or remember. Keeping the user hands and head orientation free is our key design element as it preserves the ability to interact with either the virtual or the real world.

3 Motions

3.1 Design

We start by defining as many body motions as possible that follow two principal criterias. First, the user should not need her hand, eyes nor head rotation to perform the navigation interaction. Second, the motions should be easy to perform, understand and not tiring. We end up with seven different motions, illustrated in Fig. 2 and classified into two groups: the ones that behave in the sagittal plane (Fig. 3a), i.e. bend the bust (Fig. 2b) and perform a step forward/backward (Fig. 2a). And, the ones that behave in the coronal plane or around the vertical body axis (Fig. 3a) i.e., lean the bust, translate the hips, bend the knees and rotate the shoulders or hips (respectively Fig. 2g,e,b,f,d). The step motion (Fig. 2a) is part of both groups as it behaves in the coronal plane and makes a distinction between right and left.

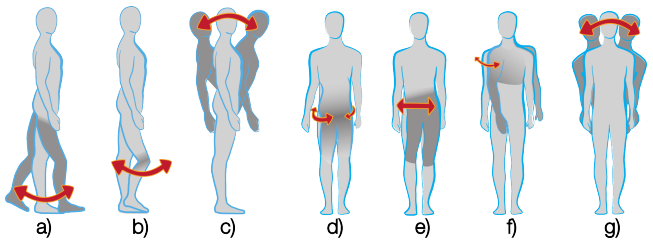

Figure 2: **Designed user motions,** a) do a step (the user just puts one foot forward or backward) b) bend knees c) bend bust d) rotate hips e) translate hips f) rotate shoulders g) lean bust.

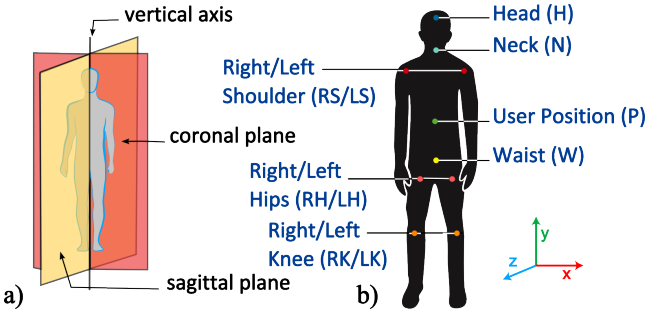

Figure 3: a) plane metaphor b) tracked points on the user

3.2 Computation

We define all our motions by measuring angles on the user body. They are computed on a set of tracked body points, captured at each timestep using an RGB-D camera located under the display (see Fig. 3b). More precisely, we measure the angles made by the body components at the current position w.r.t. a reference pose captured once by the RGB-D sensor at the beginning of the session. This makes our system adaptive to the initial pose and robust to the user morphology (e.g., height). Moreover, we couple the motion amplitude and the virtual velocity to ensure a proper speed control. In practice, we use a set of vectors defined over the tracked body key points (see Fig. 3b) to compute our angles. For the *shoulder rotation* (resp. *hips rotation*), the vector spans the two shoulders (resp. hips) positions **LS** and **RS** (resp. **LH** and **RH**). For the *lean bust* motion, the neck **N** and waist position **W** are used instead. For the *bend bust* motion, the vector goes from the user position **P** to the user head **H**, and we compute the angle in the z axis. For the *hips translation*, the angle from the user position **P** to the left hip **LH** is compared with the one from the user position **P** to the right hip **RH** in the x axis. For the *bend knee* motion, we use a vector that goes from the left knee **LK** to the right knee **RK**. Finally, to compute the *step* motion, we use the same angle than for the *bend knee* motion and we also compare the sum of the knee positions between the rest and current poses in the z axis to determine if the user goes forward or backward. We choose knees over feet because they are more likely to be inside the sensor frustum.

3.3 Assumptions

To select the aforementionned motions, we define several assumptions ground navigation shall verify to be easy to use:

Uncorrelated body parts: some body parts are easier to move in an uncorrelated way than others. For instance, it is difficult to dissociate the shoulders rotation from the hips one. Thus, the two distinct actions of the ground navigation (walking and rotating the view) should use dissociated motions and body parts.

Correlation between Virtual/and Real motions: having a good correlation between the motions performed in the real world and their effects in the virtual world helps the user to understand, remember and perform the interaction. Therefore, we consider that motions in the sagittal plane are better suited to walk, whereas motions in the coronal plane or around the vertical body axis are better suited to rotate the view.

Lazy navigation: we link the motion amplitude to the virtual speed in order to have an accurate and lazy interaction. We believe the user needs a comfortable rest pose where no interactions are happening, and at the same time "lazy" motions to navigate in the scene. Thus a good tradeoff has to be found between motions amplitude, virtual speed and rest pose.

Secondary action: ground navigation is a basic interaction, but a fully operational system may require the user to do other things and we aim at preserving the ability to perform "secondary actions" while navigating. Such actions can be either "virtual", having impact in the virtual environment (e.g., selecting, grabbing or moving 3D objects) or "real", having an effect in the real world (e.g., pointing something to someone or carrying a real object).

4 System Architecture

We provide a flexible design of our system by dividing our implementation into three main blocks: *motion receptors*, *transfer functions* and *actuators* (see Fig. 4). This allows to easily try, plug, configure, or disconnect user motions from the virtual camera. We also expose several parameters that are easy to understand and adjust, all of them being readily edited in a specific configuration file.

4.1 Motion receptors

The motion receptor block captures the motions made by the user. Using the 3D points captured by the RGB-D camera, we compute our angles between the reference and current poses (see Sec. 3.2) before normalizing them using a specific range. For each motion, the range defines the largest possible angle. Finally, we output a value between 0 and 1.

4.2 Transfer function

Based on the motion receptor angle value x, we apply a transfer function f to connect the user interaction with the virtual camera motion. This intermediate remapping procedure allows to easily plug different user movements while keeping a uniform camera motion. The output of this block is a value between 0 and 1, computed as follow:

$$f(x) = \begin{cases} 0 & \text{if} \quad x \in [0, \alpha[\\ (\frac{x-\alpha}{1-\alpha})^{\beta} & \text{if} \quad x \in [\alpha, 1] \end{cases}$$

where α controls the beginning of the motion effect (i.e., negligeable values are ignored) and β controls the slope of the function (see Fig. 4). The parameter α has a great impact on the rest pose position and allows trading-off a comfortable rest pose for small enough motions to control the navigation.

4.3 Actuator

Given the resulting normalized modulated value, we use it to control the walk (resp. rotation) speed in the virtual scene. This value is multiplied with the application-dependent maximum speed (resp. maximum rotation speed) parameter and directly used in the 3D engine camera primitives (`Move()` and `Rotate()` functionalities).

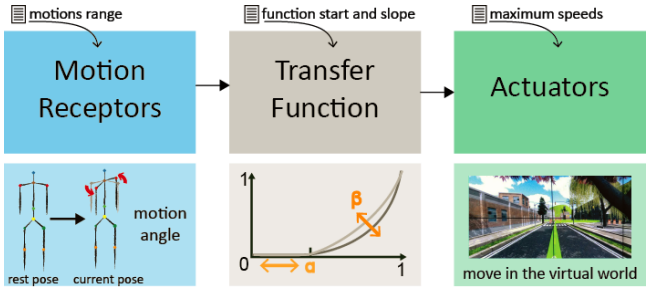

Figure 4: Architecture design.

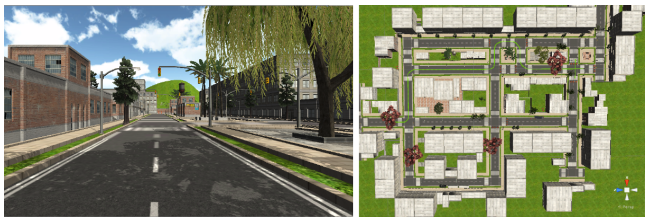

Figure 5: **Virtual Scene used in our experiment**. Left: live user view, right: top view of the scene.

5 Motion Analysis

With our architecture in hand, we analyze the resulting navigation to validate or discard our assumptions (see Sec.3.3), find out what are the best set of motions for ground navigation and outline general advices for ground navigation using mid-air 3D interactions.

5.1 System Setting

We perform our user studies using an immersive widescreen display ($4.0 : 1.15m$) made of 8 high-resolution screens (7680×2160 pixels). We use a *Microsoft Kinect* as the RGB-D sensor and the *Zigfu SDK* [15] to capture and process the tracked user body points. Our test scenario was generated using the *Unity3D* game engine. In all our experiments (i.e., pilot and user study), we use a realistic 3D scene representing a virtual city (see Fig. 5). The user initially stands far from the display (about $3.0m$) to have a better field of view. Moreover, a menu allows selecting the motions inside the application at runtime. During the initialization stage, we display the two motions currently available to the user (see Fig. 1) while the system is capturing her rest pose (no T-pose required) so that the system is as self-discoverable as possible.

We first perform a pilot user study to evaluate our assumptions and determine a first ranking of our selected motions (Sec. 5.2). Then, we keep only the best set of interactions and perform a formal user study (Sec. 5.3)

5.2 Pilot User Study

5.2.1 Procedure

We denote a pair of interactions as [V: *rotate shoulders* - W: *bend the knee*] where *V* stands for rotating the view and *W* for walking. We have 7 available motions that can be used either for walking or rotating the view, as described in Sec. 3. Therefore, as 7 sets of same motions cannot be performed for the two distinct actions, we end up with 42 sets of possible interactions. From this, we discard 6 sets of motions that were judged too much correlated to be doable: [V: *rotate the hips* - W: *translate the hips*], [V: *lean the bust* - W: *bend the bust*], [V: *bend the knee* - W: *step*] and their inverses i.e.: [V: *translate the hips* - W: *rotate the hips*], [V: *bend the bust* - W: *lean the bust*], [V: *steps* - W: *bend the knee*] (see Fig. 6).

Since the number of potential interactions is significant, we conducted a qualitative pilot study with 30 users (20 males and 10 females). Only 5 participants had already used mid-air devices to navigate in a virtual environment before. We had 6 groups of users where all users inside one group are performing the 6 same interactions. Each user performed 6 sets of motions: 3 pairs of interactions and their opposites (e.g., [V: *rotate shoulders* - W: *bend the knee*] and opposite set [V: *bend the knee* - W: *rotate shoulders*]).

5.2.2 Tasks

We designed two different tasks: first, the user discovers the motions and their effects and she can freely navigate in the virtual city. Then, when the user feels comfortable with the set of motions, we ask her to follow a virtual path displayed in the scene. As long as she is close enough, the path is green; if she goes too far, it becomes red. We repeat these two actions for our 6 different sets. We ask the user to think-aloud, and allow her to skip the path actions if she does not feel comfortable enough with the current interaction pair. Finally the user has to fill a questionnaire to give us general feedbacks on the interactions tried.

5.2.3 Results

Assumptions validation: The pilot user study allows us to validate some of our assumptions (Sec. 3.3). First, using correlated body parts to perform different actions is clearly difficult for the user. As shown in Fig. 6, users did not manage to finish the path when the walk and view interactions used too correlated body parts (e.g., *rotate the shoulders/rotate the hips*, and *translate hips/lean the bust*). Moreover people were usually able to better synchronize their actions (i.e., turning the view while walking) when the two motions were only thinly correlated. Second, having a similar correlation between real and virtual movement appeared easier. The set of motions with opposite correlation (i.e. a motion in the sagital plane to rotate the view, and a motion in the coronal plane to walk) were more difficult to perform: users report being "confused", and feel "unnatural", some users talked about "a coordination game" where it is tough to remember the good interaction, and they tend to forget the virtual impact of the interaction quickly. On the contrary, users generally need less time to understand and remember a good coordination interaction, they report them as "natural" and "easy to remember".

Favorite interactions: We ask users to rate each motions they did for the two interactions on a 5-point-Likert scale (e.g. from 1:very poor motion to 5: very good motion). We compute the Kruskal-Willis one-way ANOVA test on the obtained results and find out that there is no significant difference between the results of all the user groups for all the motions ($p = 0.05$), except for the *walk* interaction with *shoulder rotation*. This motion is significantly different among user groups ($p = 0.019$).

Finally, the favorite motions to turn the view are *rotate the shoulders* (3.24) and *lean the bust* (2.12), the other motions receiving bad scores such as, for example, (1.04) for *bend the bust*, (2.01) for *bend the knees*, and (1.35) for *do a step*. The favorite motions to walk are *bend the bust* (3.00), *bend the knees* (2.75), and *do a step* (3.42), while the others received bad scores with, for instance, (1.2) for *rotate the hips*, and (1.50) for *lean the bust*.

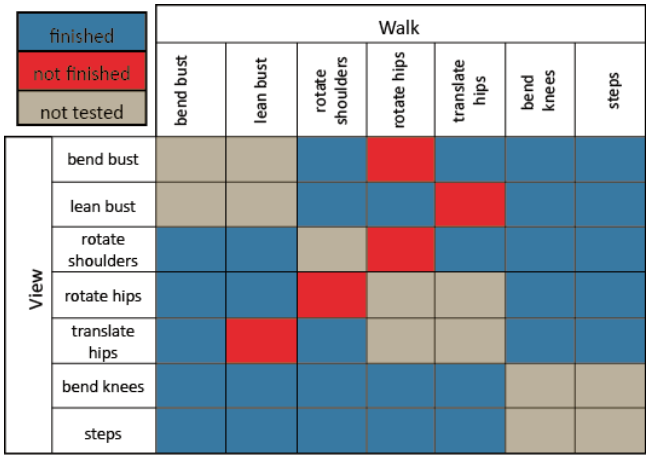

Figure 6: **Pilot user study**: motions sets.

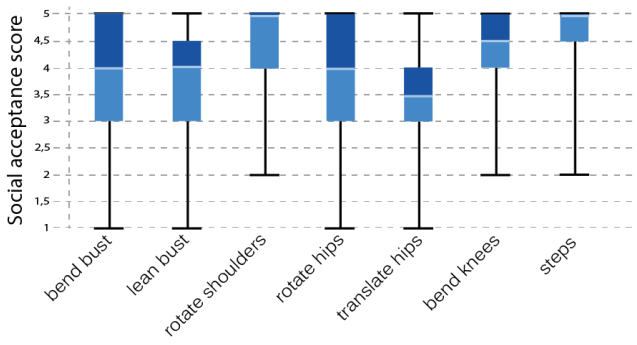

Figure 7: **Pilot user study**: users rating the motions on a 5-point-Likert scale for the question "Would you feel comfortable performing this action in public?"

Social acceptance: We ask users to tell if each motion is socially acceptable or not, with the same 5-point-Likert rating scale (see. Fig 7). We compute the Kruskal-Willis one-way ANOVA test and find out that there is no significant difference between the user groups ($p > 0.05$). The result goes from an average of (3.5) for the *translate hips* motion to an average of (4.42) and (4.52) for respectively the *bend knee* and the *rotate shoulders* motions. Therefore, we can state that all our motions are socially acceptable by most users i.e., they will feel comfortable performing them in public.

5.3 User Study

5.3.1 Procedure

Based on our pilot user study, we restricted the number of available interactions to perform a quantitative study on the best sets of actions. Following the users answers, we decided to keep only the three favorite motions to rotate the view and to walk. Moreover, these motions are consistent with our "correlation between virtual and real motions" assumption and are considered as socially acceptable. For our user study, we gathered 10 users (5 men, 5 women) aged from 20 to 28 (mean 24.5 years old) that did not participate in the pilot user study and were new to ground-navigation using mid-air devices. All the users tried the following 9 interaction possibilities in random order:

a) V: *rotate the shoulder* - W: *bend the bust*,
b) V: *rotate the shoulder* - W: *bend the knee*,
c) V: *rotate the shoulder* - W: *do a step*,
d) V: *rotate the hip* - W: *bend the bust*,
e) V: *rotate the hip* - W: *bend the knee*,
f) V: *rotate the hip* - W: *do a step*,
g) V: *lean the bust* - W: *bend the bust*,
h) V: *lean the bust* - W: *bend the knee*,
i) V: *lean the bust* - W: *do a step*.

5.3.2 Tasks

We asked each user to first get used to the current interaction pair by following a short path. During this preliminary task, the user had only to move forward, backward and to turn on the left or on the right. In the second task, we asked her to follow a path in the city as naturally as possible i.e., telling her to be as accurate, fast, and lazy as possible. Again, the path remained green as long as the user was close enough, red otherwise. For both tasks, the user was doing a secondary action while navigating in the scene, by either holding a coffee mug in her hand, carrying a grocery bag, or a backpack. After each interaction, we asked the user to rate it between 1 (very poor) to 7 (very good) for the following properties: *understandable, comfortable, easy-to-use, not tiring, accurate, secondary action doable, synchronization between walking and turning, not error prone*.

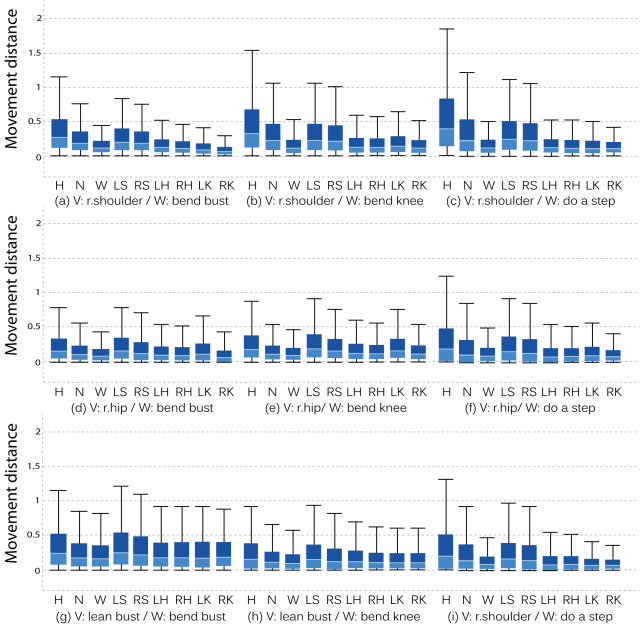

Figure 8: Average user motion for each tracked component and each interaction: head (H), neck (N), waist (W), left shoulder (LS), right shoulder (RS), left hip (LH), right hip (RH), left knee (LK), and right knee (RK).

5.3.3 Results

Movement: For each interaction, the 3D positions of the tracked user points were recorded to analyze the user motion and to get an idea of the *laziness* of each interaction; i.e. a given motion is less tiring if the user only performs small movements. Our application runs at 30Hz which means we have 30 values of all the user tracked points per second. We compute the movement made by each tracked point p_i ($i \in [1..10]$ as mentioned in Sec.3.2) for each user and each interaction, as the sum of the differences between its positions at runtime t and $t+1$ for the all user sequences:

$$m_p^i = \sum_{t_0}^{t_n} (p_{t+1}^i - p_t^i)^2$$

where t_0 and t_n are respectively the task starting and ending timestep. Fig. 8 shows a box plot of the total movement performed for each interaction and for each tracked position. The distance measured is in the normalized scale of the Microsoft Kinect sensor coordinate. The *step motion* (Fig.8 third column) generates slightly more shoulders and head movements than others. We assume this is due to the fact that the whole user body is moving while she is doing a step forward or backward. The *rotate shoulders* and *rotate hips* motions (first and second rows) have a very similar pattern. This concurs with the observation that users do not really dissociate the hips and shoulders rotation (i.e., they rotate the all bust).

To evaluate the amount of movement required by each interaction, we sum the movement of all the track points for one interaction ($motion\ quantity = \sum_0^{10} m_p^i$). Then, we perform a one-way ANOVA Kruskal-Wallis test and do not find any significant difference between the interactions ($p = 0.48$). Consequently, we cannot conclude that some motions require globally more movement than others. Therefore, to establish if some motions are more lazy than others, we rely on the user questionnaire only.

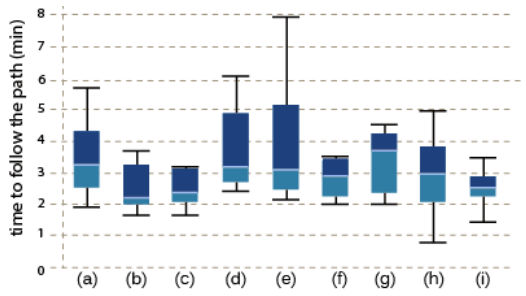

Figure 9: Boxplot representing the quartiles of time spent to follow the path (second task) for each interaction.

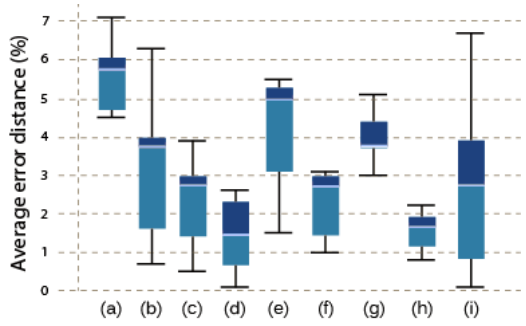

Figure 10: Percentage of average error distance of in-the-scene user position in the path following task.

Time: For each trial, the time needed to follow the path is recorded for each user and each interaction. The average time for each interaction is 3.42, 2.80, 2.89, 3.68, 3.95, 2.82, 3.90, 2.87, and 2.52 minutes, respectively. To analyze the difference among the interactions, we apply again the one-way ANOVA Kruskal-Wallis test and do not detect any significant difference ($p = 0.16$). We also apply the Wilcoxon-Signed-Ranks paired t-test to detect the significant difference ($p < 0.05$) between each pair of interactions. We find out that *a/b* ($p = 0.013$), *a/i* ($p = 0.037$), *b/d* ($p = 0.01$), *b/e* ($p = 0.01$), *c/e* ($p = 0.02$), *d/f* ($p = 0.01$), *d/i* ($p = 0.02$), *e/d* ($p = 0.03$), *e/i* ($p = 0.04$) and *d/i* ($p = 0.04$) are not significantly different. Fig. 9 shows the median value with quartile (25%:75%) of time to reach the goal of the "path follow" task. The time variation among users is different depending on the interactions. For instance, the interactions involving the *step* motion, i.e. *c, f* and *i* seems to have a smaller variance among users than others, whereas the interaction *e* results in an important time variation among users.

Accuracy: For each trial, we collected in-the-scene user positions (i.e., the position of the user avatar in the scene coordinate system for each timestep) and measured the accuracy of the user to follow the path. To do so, we compute the average of the squared distance from the user avatar to the path and we express this value as a percentage of the full track length. Fig. 10 depicts the results through their quartiles. The more accurate motions are *d* [V: *rotate hips* - W: *bend bust*] with (1.51%) of error in average and *h [V: lean bust - W: bend knees]* (1.77%), whereas the less accurate motions are *a* [V: *rotate shoulders* - W: *bend bust*] (5.72%) and *e* [V: *rotate hips* - W: *bend knees*] (5.02%).

Subjective Questionnaire: Fig. 12 shows the results of user preference (in 7-point-Likert scale) obtained for the 9 interactions and for each criteria. We confirm that our subjective questionnaire has a good reliability using the Cronbash's alpha test

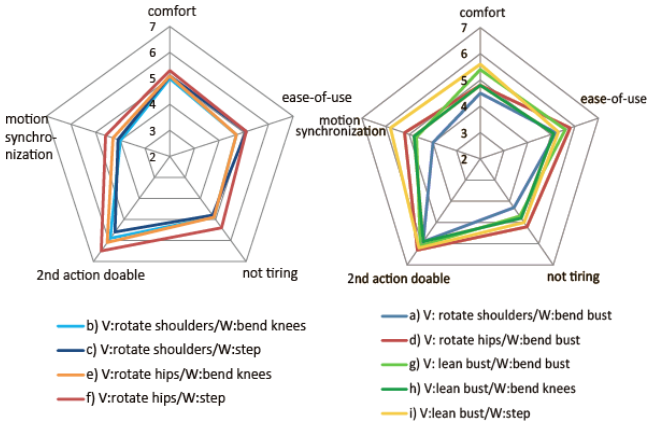

Figure 11: User ranking for the criteria with significant differences among interactions.

($\alpha = 0.724$). Overall, all the interactions were better ranked for *"understandable"* (avg. 6.56), and the *"secondary action doable"* (avg. 6.06) and they were moderately ranked (avg. 4.61) for *"the synchronization between walking and turning"* property. We analyze the user preferences in each criterion using the two-way ANOVA Friedman-Test, finding a significant difference ($p < 0.05$) between the 9 interaction sets for the following criterion: *comfortable* ($p = 0.38$), *ease-of-use* ($p = 0.12$), *not tiring* ($p = 0.52$), *secondary action doable* ($p = 0.81$), and *synchronization between walking and turning* ($p = 0.09$). As we can see on Fig. 11 (left), the interactions *b*, *c*, *e* and *f* give similar results. This supports the idea that users have a tendency to do the same gesture to rotate the hips and the shoulders i.e., in all cases they rotate the whole bust. However the angles we capture are different, especially the motion range for the *rotate hips* motions, which is smaller than the motion range for the *rotate shoulders* motions. This may explain why users rate the interactions with *rotate hips* slightly higher than the ones with *rotate shoulders* for the *not tiring* and *second action doable* criteria. The right part of the Fig. 11 shows a comparison of the other motions. We can see that the *i* and *d* motions ([V: *lean the bust* - W: *bend the knees*] and [V: *rotate hips* - W: *bend the bust*]) are overall better ranked than others, whereas the motion *i* ([V: *rotate shoulders* - W: *bend the bust*]) receive overall bad scores.

Secondary action: During the study, we asked users to perform a secondary action while navigating, it could either be: holding a coffee mug in her hand, carrying a grocery bag, or a backpack. During the user study, we observe that these actions could have some drawbacks on the interactions. First, when the user is carrying a grocery bag in her hand, the bag can interfere with the tracked body points (i.e., it moves in front of the knees and it is wrongly detected as a body point by the RGB-D sensor). This phenomenon appears more specifically when the user is rotating the shoulders or hips. And it especially has a negative impact when the rotation is coupled with the *bend knees* motion. Second, when the user is carrying a backpack she sometimes rotates her shoulders too much, the backpack then becomes visible by the RGB-D sensor and is wrongly understood as one of the shoulders. As a conclusion, bust rotations should be avoided or have a smaller range in order to avoid interferences between the navigation and some secondary actions.

6 Discussion

Based on our experiments, we now highligh several rules which appear as mandatory to develop good mid-air interactions for ground

Figure 12: Subjective preference for each interaction criterion.

navigation. Starting from a set of 7 general motions that obey to our initial conditions (no need for arm gestures or head/eye rotation), our pilot user study allowed us to restrict our approach to only 3 motions to rotate the view and 3 motions to walk. It also validates some of our assumptions. First, to be more natural, the motions in the virtual and real worlds should behave in correlated human body planes. Second, motions for distinct interactions (eg. walking and turning) should use uncorrelated body parts. Third, all the motions we kept are socially acceptable.

In the second study, we acquire a deeper understanding on the 9 best sets of motions, by measuring accuracy, time and amount of movements performed by the users. We also performed a quantitative analysis by asking the users to rate the motions on several properties. We did not clearly find a single interaction that would globally stands out from the others. However, several comments can be made from this study. First, users do not make a significative difference between rotating the hips or shoulders and only a single motion – "rotate the bust" – should be kept in a practical implementation. Rotating and leaning the bust seem to be the best motions to rotate the view. Rotating the bust appears less tiring than leaning it, but it may interfere with secondary actions. For the walk interaction, stepping, bending knees and bending the bust all have their advantages and drawbacks. While bending the bust is easy to control, it is more tiring than others. Stepping is a motion that is really easy to understand and remember but, as a discrete action, it makes more difficult to smoothly control the speed. Finally the bend knees motion is not tiring but slightly less natural: users have to remember which knee go forward.

7 Conclusion & Future Work

We have proposed a complete system for interactive mid-air ground navigation which is suitable for a large collection of applications. Our system can use several alternative body motions to control the virtual walk-through and lets critical body parts (hands, arms, head and eyes) free to perform other tasks. We support our method with a complete study that can help application developers to select a particular interaction modality for ground navigation. Although no ideal pair of interaction seems to emerge, we report a list of advices for designing a particular application scenario.

So far, we used the pilot user study to set our parameters (e.g., transfer function slope and start, maximum speed and motion ranges) as efficiently as possible, performing the main user study with a fixed set of parameters. As future work, we plan to deeper analyze the effect of our parameters on the navigation. In particular, the motion range is user-dependent and we plan to make it user-adaptive, at the potential cost of a longer initialization step which may limit applicability to public spaces.

As the analysis of the amount of performed motions for each interaction does not highlight significant differences between the interactions, we rely only on the user questionnaire to evaluate the "lazyness" of a motion. However, a deeper understanding of the interaction fatigue could be reached by building upon the work of Hincapié-Ramos et al. [16].

Last, as there is no general agreement on one best pair of motions, another direction for future work would be to design a data-driven system, learning from the user motions to adjust the interactions while she is navigating the scene. In this case, the interaction could be expressed as a weighting sum of canonical motions, optimizing the weights depending on the performed movements.

Acknowledgements This work has been supported by the JSPS KAKENHI Grant Number 25540083, the European Commission under contracts FP7-323567 HARVEST4D and FP7-287723 REVERIE and the ANR iSpace&Time project.

REFERENCES

[1] D. Bowman, S. Coquillart, B. Froehlich, M. Hirose, Y. Kitamura, K. Kiyokawa, and W. Stuerzlinger. 3d user interfaces: New directions and perspectives. *Computer Graphics and Applications, IEEE*, 28(6):20–36, Nov. 2008.

[2] J. Jankowski, T. Hulin, and M. Hachet. A study of street-level navigation techniques in 3d digital cities on mobile touch devices. In *3D User Interfaces (3DUI), 2014 IEEE Symposium on*, pages 35–38, Mar. 2014.

[3] C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the cave. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '93, pages 135–142, New York, NY, USA, 1993. ACM.

[4] T. Cakmak and H. Hager. Cyberith virtualizer: a locomotion device for virtual reality. In *ACM SIGGRAPH 2014 Emerging Technologies*, page 6. ACM, 2014.

[5] S. Ikeda, T. Sato, M. Kanbara, and N. Yokoya. An immersive telepresence system with a locomotion interface using high-resolution omnidirectional movies. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 396–399 Vol.4, Aug 2004.

[6] L. Terziman, M. Marchal, M. Emily, F. Multon, B. Arnaldi, and A. Lécuyer. Shake-your-head: Revisiting walking-in-place for desktop virtual reality. In *Proceedings of the 17th ACM Symposium on Virtual Reality Software and Technology*, VRST '10, pages 27–34, New York, NY, USA, 2010. ACM.

[7] V. K. Adhikarla, P. Wozniak, A. Barsi, D. Singhal, P. T. Kovács, and T. Balogh. Freehand interaction with large-scale 3d map data. In *3DTV-Conference(3DTV-CON), 2014*, pages 1–4. IEEE, 2014.

[8] M. Nabiyouni, B. Laha, and D. A. Bowman. Poster: Designing effective travel techniques with bare-hand interaction. In *3DUI*, pages 139–140, 2014.

[9] A. Simeone, E. Velloso, J. Alexander, and H. Gellersen. Feet movement in desktop 3d interaction. In *3DUI*, pages 71–74, 2014.

[10] F. Daiber, J. Schöning, and A. Krüger. Whole body interaction with geospatial data. In *Smart Graphics*, pages 81–92. Springer, 2009.

[11] J. Schöning, F. Daiber, A. Krüger, and M. Rohs. Using hands and feet to navigate and manipulate spatial data. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 4663–4668. ACM, 2009.

[12] J. Pettré, O. Siret, M. Marchal, J.-B. de la Rivire, and A. Lécuyer. Joyman: An immersive and entertaining interface for virtual locomotion. In *SIGGRAPH Asia 2011 Emerging Technologies*, SA '11, pages 22:1–22:1, New York, NY, USA, 2011. ACM.

[13] G. Ren, C. Li, E. O'Neill, and P. Willis. 3d freehand gestural navigation for interactive public displays. *CG&A*, 33(2):47–55, 2013.

[14] M. Roupé, P. Bosch-Sijtsema, and M. Johansson. Interactive navigation interface for virtual reality using the human body. *Computers, Environment and Urban Systems*, 43(0):42 – 50, 2014.

[15] A. Hirsch, R. Shenberg, S. Zippel, T. Blackman, and B. Tucker. Zigfu development kit for unity3d.

[16] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: A metric to quantify arm fatigue of mid-air interactions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 1063–1072. ACM, 2014.