Evin, Inan; Pesola, Toni; Kaos, Maximus; Takala, Tuukka M.; Hämäläinen, Perttu

3PP-R: Enabling Natural Movement in 3rd Person Virtual Reality

Published in:
CHI PLAY 2020 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play

DOI:
10.1145/3410404.3414239

Published: 02/11/2020

Please cite the original version:

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.
3PP-R: Enabling Natural Movement in 3rd Person Virtual Reality

Inan Evin
inan.evin@aalto.fi
Aalto University

Toni Pesola
toni.pesola@aalto.fi
Aalto University

Maximus D. Kaos
maximus.kaos@aalto.fi
Aalto University

Tuukka M. Takala
tuukkan.takala@aalto.fi
Aalto University & Waseda University

Perttu Hämäläinen
perttu.hamalainen@aalto.fi
Aalto University

Figure 1: 3PP-R allows the user to interact in 3rd-Person Perspective (3PP) using natural movements, including body rotation without losing sight of the avatar. A virtual display such as a 3D miniature world model hovers in air in front of the user, showing a 3rd-person avatar. When the user turns, the display orbits around the user but does not rotate except for the avatar. From the user’s perspective, the display appears fixed in the field of vision, and the world rotates around the avatar.

ABSTRACT
We propose 3PP-R, a novel Virtual Reality display and interaction technique that allows natural movement in 3rd-person perspective (3PP), including body rotation without losing sight of the avatar. A virtual display such as a World-in-Miniature model orbits around the user when the user turns, but does not rotate except for the user’s avatar. From the user’s perspective, the display appears fixed in the field of vision, while the world rotates around the avatar. 3PP-R combines the strengths of 3PP and 1st-person perspective (1PP): Similar to 1PP, it allows interacting with rich natural movements, while also reaping the benefits of 3PP, i.e., superior spatial awareness and animating the avatar without nauseating viewpoint movement, e.g., for joystick-controlled locomotion. We test 3PP-R in a maze navigation study, which indicates considerably less cybersickness in 3PP-R than in 1PP. We also demonstrate 3PP-R in dynamic game interaction including running, jumping, swinging on bars, and martial arts.

CCS CONCEPTS
• Computing methodologies → Virtual reality; • Human-centered computing → Interaction techniques.

KEYWORDS
Virtual reality display; 3rd person perspective; Virtual camera design

ACM Reference Format:

1 INTRODUCTION
A fundamental problem of Virtual Reality (VR) is how to seamlessly combine three features: 1) interaction using natural body movements, 2) the ability to navigate large virtual spaces, and 3) keeping cybersickness to a minimum. Using a 1st person perspective (1PP) and one-to-one viewpoint tracking minimizes cybersickness and allows natural interaction, but at the same time, it prevents navigating virtual spaces larger than the real interaction space. To solve the
problem, various locomotion approaches have been developed, but they all have their drawbacks. For example, moving using a joystick can be nauseating [44], point-and-teleport or world-in-miniature (WiM) [66] approaches can break interaction flow or be unrealistic during interactions such as martial arts fighting. Walk-in-Place (WiP) locomotion can feel more immersive than teleporting, but may cause discomfort [13], and the required movement detection and classification inevitably adds latency and may interfere with detecting other movements such as jumping.

Alternatively, some VR experiences utilize 3rd person perspective (3PP) where the user views the avatar from outside. This is also common in non-VR movement-based games, e.g., for the Kinect [71]. 3PP provides multiple benefits. First, 3PP enables controlling the avatar using a joystick without inducing a visual-vestibular sensory conflict, a major cause of motion sickness [1]. Seeing one’s avatar from outside also allows for more artistic freedom, e.g., in animation design. The ability to freely animate the avatar also overcomes the need for tactile feedback, e.g., in making impacts knock back the avatar during a melee. In 1PP, such reactions would deprive the user from control of the viewpoint, which is strongly advised against in order to minimize nausea [48, 51].

3PP interaction is not without problems, however. In particular, the need to watch the avatar means the player cannot freely turn around, which limits interaction using natural body movement. This is not ideal, especially as modern wireless VR headsets such as the Oculus Quest allow free body rotation without the user getting tangled up with wires.

To solve both the 3PP rotation problem and the 1PP locomotion problem, we propose and evaluate a novel VR display and interaction technique we call 3PP-R, illustrated in Figure 1. In 3PP-R, a virtual display such as a miniature 3D world model hovers in air in front of the user, showing a 3rd person avatar of the user. When the user turns, the display orbits around the user but does not rotate except for the avatar. From the user’s perspective, the display appears fixed in the field of vision, and the 3D world rotates around the avatar. To prevent the feeling of self-motion and minimize cybersickness, both the user and the 3PP-R display are embedded in a non-moving 1st person reference world, and the display volume is limited using a cutout shader.

This paper makes the following contributions:

- A demonstration of 3PP-R in rich, dynamic VR game interactions (Figure 5, supplemental video), including joystick-controlled running, exaggerated jumping between platforms, physically simulated swinging on bars, and martial arts fighting where the user dodges and counters attacks with full-body movements, all in a small real-world space. In 1PP, this would inevitably cause cybersickness due to the deviation from natural one-to-one viewpoint tracking required by the joystick control and physics simulation of the avatar.

- Our 3PP-R prototypes work on low-cost VR hardware. The user study was conducted and the supplemental video was captured on an Oculus Quest mobile VR headset.

2 BACKGROUND AND RELATED WORK

Display and interaction approaches other than natural 1PP are primarily motivated by the limited space available for the user in the real world, which makes it impossible to walk or run around large virtual spaces. To address the problem, various VR locomotion and navigation approaches have been developed. There is also some, albeit much more limited, research on 3PP interaction and approaches combining 1PP and 3PP. Below, we review these topics briefly.

We also discuss cybersickness, as it is the main obstacle for using simple and obvious locomotion controls such as a joystick. Following [3, 26], we use the term cybersickness for motion sickness induced by VR. A related term is Visually Induced Motion Sickness (VIMS), but Arcioni et al. [3] argue that cybersickness can also be non-visual or multisensory in origin.

2.1 Locomotion Techniques and Cybersickness

A multitude of VR locomotion techniques exists. The following discussion utilizes a categorization into continuous virtual movement, discrete transitions, and manipulations of natural locomotion. However, this is only scratching the surface, and for more comprehensive reviews, the reader is referred to LaViola Jr et al. [46, p. 318] and Boletis [12].

2.1.1 Continuous Virtual Movement. The joystick-controlled locomotion of our 3PP-R prototypes belongs to a larger class of steering locomotion [46, p. 339]. In steering locomotion, virtual movement is commonly generated by pressing a button or pushing a joystick, while the forward direction for locomotion is usually determined by the direction of the user’s gaze or by pointing. Other alternatives for generating virtual movement include walking-in-place [64] and arm-swinging gestures [77].

Locomotion techniques differ in the amount of cybersickness that they cause, and steering locomotion is known to be a serious offender in this regard [17, 44]. The oldest and most prominent cybersickness theory — the sensory conflict theory — proposes that sensory conflicts between visual and vestibular cues induce cybersickness [45]. Therefore many approaches for reducing cybersickness focus on minimizing optical flow that is unrelated to the user’s physical motion, particularly when it comes to acceleration. In the case of steering locomotion, such cybersickness mitigation approaches include field of view restriction [24] and Independent Visual Background (IVB) that matches the vestibular cues [21]. Wu and Rosenberg [79] utilized both of the aforementioned approaches.

1Hence the “R” in 3PP-R, denoting the ability to rotate one’s body without losing sight of the 3PP world and avatar.
We extend the approach to 3PP, while adding the ability to turn around without losing sight of the 3rd person avatar.

2.1.2 Discrete Transitions. Instead of continuous virtual movement, one can also use discrete transitions in the form of teleportation [16] and its “dash” variants [9, 14], where the user viewpoint is quickly translated to the target location. Instantaneous teleportation induces less cybersickness than steering [17, 25, 44], but this comes at the expense of spatial orientation [6, 15].

Clifton and Palmisano [18] noted large individual differences between participants of their study, where they compared steering and teleport locomotion in terms of cybersickness and presence. They concluded that it may be difficult to devise a universally favored locomotion method. This has not gone unnoticed by VR developers either; many popular VR games like Arizona Sunshine [76], Half-Life: Alyx [74], and Raw Data [70] provide both steering and teleport locomotion options for players to choose from.

One can also trigger discrete transitions using the World-in-Miniature (WiM) interaction metaphor [66]. WiM utilizes one or more minismaps of the virtual environment, which one can hold and orient in one’s hand. If the minimap is focused on the area where the user is located, this location on the minimap contains a human figure that indicates the user’s position and orientation. Locomotion can be achieved in WiM by selecting the human figure and dragging it to another location in the minimap, after which the user experiences an animated transition in 1PP from the current location to the target location. Wingrave et al. [78] extended WiM by allowing users to scale the WiMs and work at different levels of scale. In contrast to WiM, the 3PP-R miniature display is always in front of the user and the user’s avatar is always at the center of the display. Moreover, locomotion in 3PP-R does not require selection and dragging. In the context of WiMs, it should also be noted that since 3PP-R combines a miniature display with a natural-scale 1st person reference world, 3PP-R could be considered an instance of a Multi-Scale Virtual Environment (MSVE) [41].

2.1.3 Manipulating Natural Locomotion. The real-world space required by natural locomotion can be decreased by manipulating scale and orientation. In GullIVR [42], the user moves across the virtual environment by walking within the motion-tracked interaction space. The user has the ability to switch between being a giant and normal size, former of which allows traversing large areas in the virtual environment. The drawback of GullIVR is that it is not ideal for virtual environments that have ceilings. Redirected walking techniques, on the other hand, tackle the issue of limited interaction space by redirecting the user away from its boundaries. This redirection can be achieved in various manners, for example by continuously applying imperceptible rotations to the virtual world while the user is walking [35], or by leveraging change blindness and covertly rearranging the layout of the virtual environment while the user is looking away [69]. In practice, redirected walking techniques still require a relatively large interaction space, when compared to our 3PP-R and joystick locomotion.

2.2 3rd Person VR

Some VR platformer games, like Astro Bot Rescue Mission [65], Lucky’s Tale [58], and Moss [59], utilize 3PP in a manner analogous to that of traditional 3D games; the viewpoint location follows the gamepad-controlled player character from a distance, while the player’s head movements act as the primary camera control in 1PP fashion. In 3PP-R the avatar can be controlled using natural movements in addition to a gamepad or handheld VR controller, without losing sight of the avatar when turning. Also, 3PP-R uses field-of-view restriction and an independent visual background to reduce cybersickness, both of which are absent in the above-mentioned VR games.

Although having a 3rd person, bird’s eye view of an environment provides spatial awareness superior to 1PP, it should be noted that 3PP sets some limits on interaction: interaction with virtual objects is less accurate, e.g., when catching or deflecting projectiles [2, 30]. On the other hand, the problems can be mitigated by techniques such as predictive visualization of movement trajectories, or sliding the avatar towards a target while punching [2]. We implement the latter in our 3PP-R prototype. Another limitation of 3PP may be in the degree of presence [32]. Presence, a key aspect of VR experiences, denotes the feeling of being or acting in a place, regardless of where one is physically located [60]. 1PP is considered to provide the greatest feeling of presence [53, 57, 63], and 3PP inherently detaches the user and the avatar, although some studies have not found an effect of perspective on presence [20, 30].

2.3 Combining 1st and 3rd Person Perspectives

The VR game Front Defense: Heroes [23] includes a locomotion technique called V-move, where pushing a button moves the avatar in the controller’s forward direction, while the viewpoint does not move until the button is released, at which point the viewpoint teleports to the avatar position. Effectively, this means that the 1PP changes to 3PP for the duration of the locomotion. Griffin and Folmer [32] presented a similar locomotion technique, where a button can switch from 1PP to 3PP for the duration of steering locomotion. Cmentowski et al. [19] introduced a related technique, where switching to 3PP gives the user a giant’s viewpoint over the virtual environment, but leaves the avatar’s size unaffected. Common to these three locomotion techniques is that they alternate between 1PP and 3PP. This is different from 3PP-R, where the miniature display maintains a 3PP view on the avatar, while the surrounding independent visual background is simultaneously seen in 1PP. Also, in Front Defense: Heroes [23] as well as [19, 32], the user loses sight of the avatar if turning around while in the 3PP view.

2.4 Summary

Using the terminology of the approaches reviewed above, 3PP-R combines 3PP with steering locomotion, field-of-view restriction, independent visual background, and a novel WiM-style minimap that orbits the user to stay visible when the user turns. This provides considerable freedom for interaction design and can be directly applicable to games such as Boneworks [68], which uses physically based interactions that allow the world to affect the avatar’s movement. Games like Boneworks could and possibly should utilize 3PP-R: players love the interaction and the game quickly sold over 100k copies during its launch week, but the present 1PP version is divisive in how much nausea and discomfort it causes [4, 75].
3 SYSTEM

Our proposed 3PP-R approach can be summarized as follows, with more details and design rationale provided in the sections below.

- A virtual 3PP display hovers in the air in front of the user, as shown in Figure 1. The display orbits around the user, allowing 360 degree rotation. We use a World-in-Miniature 3D model, but it should also be possible to use a virtual planar 2D screen.
- The virtual display shows both the virtual world and a 3rd-person avatar. When the user turns, the avatar turns as well, but the world around the avatar does not rotate, as illustrated in Figure 1.
- The avatar mimics the user’s movements but large-scale locomotion is also enabled using a joystick.
- To minimize conflicting visual and vestibular cues caused by the joystick control, 1) the visible game world is limited by only rendering a volume around the avatar, and 2) both the user and the virtual display are embedded in a static 1PP reference world (an IVB). In our prototype, we use a floor plane with a checkerboard pattern as the reference.

To allow dynamic action gameplay and demonstrate how 3PP-R allows flexible manipulation of the user-avatar movement mapping, our 3PP-R game prototype also implements additional features detailed in Section 3.5. We exaggerate the user’s tracked locomotion speed and jump height, and increase punching reach by sliding the user towards punch targets. We also limit sideways joystick locomotion speed to encourage using body movements instead of “cheating” by using the joystick all the time.

Our 3PP-R prototypes were implemented for the mobile stand-alone Oculus Quest VR head-mounted display (HMD) [55]. The Quest’s handheld controllers feature analog joysticks and 6 degree-of-freedom tracking. We used the Unity 3D game engine [72] which provides basic tools such as inverse kinematics, avatar animation and physics simulation. Below, we also discuss 3PP-R in terms of standard 3D graphics view and model matrices, in order to allow reproducing our work without a specific game engine.

3.1 The Miniature Display

At first thought, it may seem that implementing our World-in-Miniature 3PP-R display is simply a matter of scaling. However, scaling the game world is not desirable for a number of reasons. In modern game production, one often works with a number of 3rd party 3D assets and scripts which may not work with scaling. As pointed out by [42], resizing the environment may also interfere with baked lighting. Finally, adjusting the scale dynamically during a game—together with physics simulation time step and object masses to preserve natural physics simulation behavior—can easily result in simulation glitches.

To make 3PP-R a simple drop-in replacement for traditional 1PP VR without requiring changes to game content, we do not scale the game world and only provide an illusion of miniaturization. We implement this through 1) manipulating the VR camera (the view matrix), and 2) scaling and moving the 1PP reference world around the player. It is assumed that the 1PP world is only a visual reference without physics simulation and can thus be freely moved and scaled.

In standard computer graphics terms, recall that when using homogeneous coordinates, a rendered point \( x' \) is computed from a model point \( x \) through matrix multiplication as:

\[
x' = PMx,
\]

where \( P, V, M \) are the projection, view, and model matrices, respectively. Usually, VR tools like the Oculus VR plugin handle the projection matrix, and a developer only manipulates the view and model matrices. The model matrix transforms points from local model coordinates to the world space, and the view matrix transforms from world space to camera-local coordinates.

The model matrix can be expressed as:

\[
M = \begin{bmatrix} s_m & R_m & p_m \\ 0 & 1 & 1 \end{bmatrix},
\]

where \( R_m, p_m, s_m \) are the rotation, position, and scale of the displayed 3D object. For simplicity, we assume uniform scale, i.e., scalar \( s_m \). When using a game engine like Unity, one typically manipulates rotation, position, and scale using an API and/or a graphical user interface, and the model matrix is composed by the engine under the hood.

If one considers the camera as another 3D object in the world, the view matrix can be expressed as:

\[
V = \begin{bmatrix} s_c & R_c & p_c \\ 0 & 1 & 1 \end{bmatrix}^{-1},
\]

where \( R_c, p_c, s_c \) are the camera rotation, position and scale.

The important thing to note is that usually, one does not manipulate the camera scale. However, it is useful in our case, as a large \( s_c \) makes everything appear smaller. We let the HMD control \( R_c \), treat \( s_c \) as a parameter adjusted by the developer or user, and calculate camera position as:

\[
p_c = p_a + s_c(-df_{hmd} + (ph_{hmd} - h)u_y),
\]

where \( p_a \) is the avatar’s position, \( df_{hmd} \) is the HMD’s forward vector projected on the ground plane and \( u_y \) is the global up vector (y-axis). As illustrated in Figure 2, \( d \) denotes the horizontal camera distance from the avatar, \( h \) is the desired elevation of the avatar in relation to the 1PP ground plane, and the tracked vertical HMD position is denoted by \( ph_{hmd} \). The avatar’s position \( p_a \) is controlled both by HMD movement and a handheld controller/joystick.

3.2 The 1st Person Reference World

To provide the user a feeling of natural movement and non-conflicting visual and vestibular sensations, we include a 1st person reference world surrounding both the user and the miniature display, as illustrated in Figure 1. This kind of Independent Visual Background (IVB) has been previously found to alleviate simulator sickness [21]. Since we want the IVB to appear static despite the camera manipulation, we scale it up by \( s_c \), and set its position as:

\[
p_{1st} = p_a + s_c(-df_{hmd} - hu_y - ph_{hmd} xz).\]

where \( ph_{hmd} xz \) is the HMD’s tracked horizontal position.

It should be noted that for stereographic rendering, two cameras are needed. However, VR toolkits usually handle this under the
we
For
dual
3.5
threshold
distance
user’s
of
initial
sponds
to
position.
HMD
tional
punching,
In
the
to
two.
hood,
and
defining
a
single
camera
position
is
enough,
corresponding
to
the
midpoint
between
the
two.

3.3 Calibration
In
our
Oculus
Quest
implementation,
no
3PP-R
specific
 calibration
is
needed
for
basic
locomotion
and
interaction.
The
Quest
itself
provides
HMD
tracking
and
floor
plane
estimation,
and
also
allows
the
user
to
reset
the
origin
of
the
HMD
tracking
with
a
dedicated
button.
In
our
action
demo,
we
utilize
inverse
kinematics
for
punching,
grabbing
objects,
and
crouching,
which
requires
additional
knowledge
of
the
user’s
height
and
limb
lengths.
These
are
estimated
from
a
set
of
 calibration
poses
the
user
is
instructed
to
take.

When
the
user
loads
and
starts
a
3PP-R
application,
the
initial
HMD
position
is
saved,
and
the
HMD
positions
used
in
equations
4
and
5
are
offset
such
that
the
origin
of
the
HMD
tracking
corresponds
to
the
initial
avatar
position.

3.4 Limiting
the
display
volume
Large
virtual
worlds
can
penetrate
the
user’s
body
even
when
miniaturized,
which
can
feel
disturbing.
Also,
to
prevent
feelings
of
self-motion,
moving
objects
that
take
up
a
large
portion
of
the
user’s
field
of
view
should
be
avoided
[51].
Because
of
this,
we
utilize
a
cutout
shader
that
culls
geometry
further
from
the
avatar
than
a
distance
threshold
\(t_0\).

3.5 Locomotion
Control
For
horizontal
movement,
we
implement
a
dual
approach:

- **HMD
  and
  hand
  tracking:**
The
HMD’s
horizontal
movement
is
mapped
to
avatar
locomotion
speed
with
an
exaggeration
scale
\(s_{\text{H}}\)
if
the
HMD
is
moving
faster
than
a
threshold
\(t_1\).
Movement
is
also
empowered
by
sliding
the
avatar
further
towards
targets
when
the
user
punches.
Punching
is
detected
as
hands
moving
away
from
the
shoulders
faster
than
a
threshold
\(t_p\).

- **Joystick
  control:**
The
user
can
move
the
avatar
using
the
Oculus
Quest
handheld
controller’s
joystick.
Pushing
the
joystick
slightly
forward
or
back
makes
the
avatar
move
at
walking
speed,
and
pushing
the
joystick
all
the
way
forward
or
back
makes
the
avatar
run.
Sideways
movement
speed
is
limited
by
scaling
it
with
a
parameter
\(s_s\).

Limiting
the
sideways
movement
speed
is
crucial:
With
the
limit,
joystick
control
allows
convenient
navigation
of
large
virtual
 worlds,
while
also
encouraging
the
use
of
natural
body
movement.
The
latter
is
desirable
from
the
perspective
of
presence,
engagement,
and
body
ownership
[11, 29, 62, 73].
Without
the
speed
limit,
one
very
easily
stops
rotating
or
moving
one’s
body
and
only
uses
the
joystick
for
all
movement.
We
noticed
this
in
initial
testing,
e.g.,
in
the
game
demo
where
the
user
needs
to
dodge
projectiles
by
 moving
sideways
(Figure
5A,
supplemental
video
at
00:50).
The
observation
is
in
line
with
research
that
shows
that
human
behavior
can
to
a
large
extent
be
explained
as
utility
maximization,
merging
with
limited
computational
 capabilities
[27].
The
joystick
control
 has
high
utility
and
low
cost,
in
particular
in
3PP-R,
where
it
does
not
cause
nausea,
which
explains
how
users
might
adopt
it
as
the
primary
locomotion
approach.
Research
on
(news)-based
games
has
also
found
that
players
 tend
to
minimize
movements
not
needed
for
a
task
[10].
We
also
tried
completely
disabling
joystick-based
sideways
locomotion,
but
that
made
it
difficult
to
precisely
control
locomotion
direction.

The
slide-to-target
punch
mechanism
allows
hitting
targets
with
less
walking/running
and
enables
dynamic
martial
arts
fighting
in
a
small
real-world
play
area.

For
vertical
movement,
we
employ
the
following:

- **Exaggerated
  jumping:**
If
the
player
moves
upwards
faster
than
a
threshold
\(t_j\)
and
is
on
the
ground
or
has
been
on
the
ground
during
past
200ms,
the
avatar’s
vertical
velocity
is
set
to
the
HMD’s
vertical
velocity
multiplied
by
a
scalar
\(s_j\).
Otherwise,
the
avatar’s
vertical
velocity
is
affected
by
gravity.

- **Staircases
  and
  ramps:**
The
avatar’s
animation
controller
uses
a
standard
spherically
capped
 cylinder
as
the
collision
geometry,
which
allows
it
to
slide
on
stairs
and
ramps
when
it
is
moved
horizontally.

In
summary,
we
implement
multiple
ways
of
exaggerating
and
empowering
movement.
We
scale
up
the
user’s
walking/running
speed
and
jumping
height,
and
also
increase
punching
reach
through
the
avatar
sliding.
These
reduce
the
space
needed
by
natural
movement
interaction,
and
are
also
motivated
by
research
that
has
found
exaggerated
movement
desirable
for
the
user
experience
[31, 33, 47],
and
that
users
 can
 perceive
moderately
exaggerated
movement
as
more
natural
than
fully
realistic
movement
[31].

3.6 Parameters
Table
1
summarizes
the
parameters
discussed
above
and
their
values
used
in
our
3PP-R
prototype.
The
parameter
values
were
determined
empirically
by
the
authors
during
initial
prototyping.
An
in-depth
 investigation
of
the
user
preferences
for
the
parameter
values
is
deferred
to
future
work.
The order spheres. 

Table 1: Parameters and their values used in our prototypes.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_c$</td>
<td>50</td>
<td>Camera scale</td>
</tr>
<tr>
<td>$d$</td>
<td>0.75 m</td>
<td>Camera horizontal distance from avatar</td>
</tr>
<tr>
<td>$h$</td>
<td>0.79 m</td>
<td>Avatar elevation from 1PP ground plane</td>
</tr>
<tr>
<td>$t_D$</td>
<td>0.28 m</td>
<td>Cutout shader distance threshold</td>
</tr>
<tr>
<td>$t_J$</td>
<td>2.3 m/s</td>
<td>Tracked jump speed threshold</td>
</tr>
<tr>
<td>$s_J$</td>
<td>3.84</td>
<td>Tracked jump speed scale</td>
</tr>
<tr>
<td>$s_g$</td>
<td>8.0</td>
<td>Gravity speed scale</td>
</tr>
<tr>
<td>$t_P$</td>
<td>2.2m/s</td>
<td>Tracked punch speed threshold</td>
</tr>
<tr>
<td>$t_L$</td>
<td>0.46 m/s</td>
<td>Tracked locomotion speed threshold</td>
</tr>
<tr>
<td>$s_L$</td>
<td>2.66</td>
<td>Tracked locomotion speed scale</td>
</tr>
<tr>
<td>$s_S$</td>
<td>0.33</td>
<td>Joystick sideways locomotion speed scale</td>
</tr>
</tbody>
</table>

4 EXPERIMENT 1: MAZE NAVIGATION

To verify that 3PP-R—similar to 3PP in general—offers more possibilities for avatar control without cybersickness, we conducted a single-session within-subjects study (N=38) comparing 3PP-R to 1PP. We used the joystick-controlled maze navigation task illustrated in Figure 3. The study was designed to answer six research questions:

- RQ1. Which approach causes the least cybersickness?
- RQ2. Which approach produces the strongest feeling of embodiment?
- RQ3. Which approach is the easiest to use?
- RQ4. Which approach produces the strongest feeling of competence? Feeling competent is a basic psychological need, a central intrinsic motivation factor [61], and considered essential for good user experience [34].
- RQ5. Which approach is considered more natural?
- RQ6. Which approach do participants prefer?

4.1 Task design

The participants were asked to navigate through a maze using an Oculus Quest handheld controller’s joystick, collecting 22 waypoint spheres. The task was repeated in both 1PP and 3PP-R, with the order of experimental conditions counterbalanced. The task was designed to last 1-2 minutes, which we expected to be long enough for cybersickness to manifest, but short enough to avoid excessive sickness. The maze started from an empty room where the participants could try out the locomotion, and maze completion time was measured starting from reaching the first sphere placed in a corridor leading out from the initial room.

As the amount of head rotation may affect cybersickness [8, 52, 56], we minimized head rotation differences between the experimental conditions by using maze walls so high that one could not see the next sphere in 3PP-R without turning one’s head in the correct direction, similar to 1PP.

4.2 Hypotheses

We formulated the following hypotheses for RQ1 and RQ2:

- H1: Joystick-controlled locomotion causes significantly less cybersickness in 3PP-R than in 1PP.
- H2: 1PP causes a stronger feeling of embodiment.

As detailed in Section 2, 3PP-R should reduce cybersickness due to the display volume limitation and the independent visual background. Testing H1 was still considered essential, as no previous study had tested a 3PP display that orbits around the user when the user turns; this novel visual stimuli might feel unnatural or cause unpleasant sensations.

Regarding H2, while there is some evidence to the contrary [20], first-person perspective is generally considered to offer higher embodiment [57, 63].

4.3 Exploratory Research Questions

RQ3-RQ6 are exploratory, without strong a priori hypotheses. We investigated them to inform future research and applications of 3PP-R.

Regarding RQ3 and RQ4, given the greater spatial awareness that 3PP-R provides over 1PP (Figure 3, also [30]), navigating a maze in 3PP-R could be considerably faster, which could lead to higher ratings on ease of use and competence. However, the high walls of our maze prevented participants from perceiving more navigation targets and planning further ahead in 3PP-R.

Regarding RQ5 and RQ6, there are a multitude of factors that may affect perceived naturalness and the preferred perspective. Hence, we avoided making a priori hypotheses.

4.4 Inclusion Criteria

Eligibility criteria were primarily based on the Oculus Quest Health and Safety Warnings guide [22]. We instructed people to only consent to participate when unimpaired and not suffering from pre-existing medical conditions that may be exacerbated by use of the virtual reality headset and controllers. Exclusionary impairments included being tired or needing sleep, under the influence of drugs or hung-over, currently experiencing digestive issues, under emotional stress or anxiety, or suffering from cold, flu, headaches, migraines, or earaches. Exclusionary pre-existing medical conditions included being pregnant, having pre-existing binocular vision abnormalities (such as double vision), or suffering from a heart or other serious medical condition. In addition, participants who had a history of seizures or an implanted medical device (e.g., cardiac pacemakers, hearing aids, and defibrillators) were excluded from participating.
4.5 Participants and Recruitment

38 participants were recruited from Aalto University campus using advertisements placed at building entrances. A movie ticket was given as an honorarium for participating in the study. In a demographics questionnaire, gender was asked in short answer format. 13 participants reported being female (34.2%), and 25 participants reported being male (65.8%). Participant age ranged from 15 to 44 years old ($M = 27.6$, $Mdn = 27.0$, $SD = 6.11$). Participants were asked to rate their video game experience and virtual reality experience on 5-point Likert scales. The results of these questions indicated significant video game experience, $M = 3.67$, $Mdn = 4.00$, $SD = 1.08$; and relatively low virtual reality experience, $M = 2.00$, $Mdn = 2.00$, $SD = .771$.

4.6 Outcome Measures

The study had six outcome measures: cybersickness, avatar embodiment, usability (comprising two measures: ease of use and naturalness), competence, and player preferences. We measured cybersickness using the Simulator Sickness Questionnaire [39]. The SSQ is a multi-faceted measure of sickness with factors of Nausea (e.g., stomach awareness, increased salivation, sweating), Ocullomotor (e.g., eyestrain, difficulty focusing, blurred vision), and Disorientation (e.g., dizziness, vertigo). While there is some debate over its applicability to virtual reality [67], recent findings suggest that the symptoms of cybersickness and motion sickness are not different [49]. In addition, the SSQ has stood the test of time [7]. VR-specific questionnaires have been derived directly from the SSQ [40], and recent VR studies still use the SSQ as a basis for measuring cybersickness [32, 44].

To further test cybersickness, in line with postural instability theory and prior cybersickness research [54], we used the stand on preferred leg (SOPL) test [39]. As we did not have a method readily available for measuring hip sway, we elected to use complete failure (i.e., participant falls when they must put both feet down) as the failure criteria, or a maximum of 30 seconds, whichever came first.

To measure embodiment, we use the standardized questionnaire from Gonzales-Franco and Peck [28]. The questionnaire contains 25 seven-point Likert scale questions with six subscales. After reviewing the questions, only the agency subscale was deemed applicable to our study. For instance, the body ownership subscale contains the question, “I felt as if the virtual ___ I saw when looking in the mirror was my own ___”, which was not applicable since there was no mirror in the game. The tactile sensations subscale was not applicable because there was no direct interaction with objects in the game.

For usability, including whether the VR interaction method was easy and natural to use, we modified the questionnaire proposed by McMahak et al. [50] to fit our study. This questionnaire consists of 12 VR usability questions, and while not validated, the questions were deemed highly relevant to measuring the usability of our interaction method.

To measure competence, we used the competence subscale of the Ubisoft Perceived Experience Questionnaire, a validated questionnaire consisting of five-point Likert scale questions based on Self-Determination Theory [5]. We also measured task completion time as a potential contributor to competence.

To assess player preferences, we conducted semi-structured interviews after the last questionnaire was filled out. In the approximately two-minute interviews, we asked:

1. You played through a maze in first and third person perspective. If you had to choose, which of the two navigation methods would you prefer to use in a VR action game and why?
2. What do you think are the pros and cons of each perspective?
3. Do you have any additional feedback you would like to provide?

4.7 Procedure

Prior to study commencement, consent was obtained if the participant was at least 15 years old, as per the national research ethics board recommendation of Finland. If the participant was under the age of 15, parental consent and child assent were obtained. Once deemed eligible, the participant first filled out a demographics questionnaire. Next, a sickness/nausea rating was obtained related to the participant’s current health status. In accordance with Howarth and Finch [36], this rating was measured by a single four-point Likert scale question, current sickness rating (SR), with possible responses of “No symptoms”, “Any symptoms, but no nausea”, “Mild nausea”, and “Moderate nausea”. The participant was instructed to tell the experimenter and continue the experiment if moderate nausea was selected. Otherwise, the participant then filled out the Simulator Sickness Questionnaire (SSQ) [39]. The participant then completed the first maze navigation task. Once complete, the participant again answered the SR. If “Moderate nausea” was selected, the study was discontinued. Otherwise, the participant next filled out the SSQ and performed the SOPL postural stability test [54]. The participant then completed the primary study questionnaire (PSQ), which measured avatar embodiment [28], usability [50], and competence [5]. The participant then completed the second maze navigation task. Once complete, the participant answered the SR, followed by the SSQ if “Moderate nausea” was not selected, then performed the SOPL stability test, and filled out the PSQ. Finally, the participant answered questions in a semi-structured interview regarding their preferences.

4.8 Sample Size Determination

G*Power 3 was used to conduct a priori power analysis to determine the total sample size necessary for the study. Power analysis for paired-samples t-tests showed a sample of 34 participants would be required to detect a medium effect size ($d_z = .50$) with a type one error rate 0.05, and 80% power [43].

4.9 Statistical Analyses

Statistical analyses were performed using SPSS version 26.0 [37]. All statistical tests were two-tailed and maintained a 5% confidence interval. Paired-samples t-tests were performed on avatar embodiment and UPEQ competence. Related-samples Wilcoxon signed rank tests were performed on cybersickness, usability, and task completion time, due to Shapiro-Wilk tests of normality indicating that the data was not normally distributed. For all hypothesis testing, Holm-Bonferroni adjusted p-values are reported.
Figure 4: Results comparing 3PP-R and 1PP in the maze navigation test. 3PP-R scores significantly better in the Simulator Sickness Questionnaire (SSQ), at the cost of somewhat lower embodiment. Holm-Bonferroni corrected p-values are reported.

4.10 Results

The results are summarized in Figure 4. The data provides support for both hypotheses H1 and H2, indicating that participants experienced very little cybersickness in 3PP-R, and considerably more cybersickness in 1PP, but on the other hand, 1PP offered a stronger feeling of embodiment/agency. Despite the high 1PP cybersickness scores, there was no participant dropout. There were no statistically significant differences in the other outcome measures. Participant preferences were roughly equally divided, some favoring 3PP-R because of less cybersickness, and some favoring 1PP because of greater immersion. Below, we explain the results in more detail.

Cybersickness: Results of a Shapiro-Wilk test of normality showed that SSQ data was not normally distributed. Thus, a Wilcoxon signed rank test was performed. Results of this test showed a statistically significant difference in SSQ ratings, with 1PP (M = 12.7, SD = 15.2) resulting in statistically significantly greater SSQ ratings over 3PP-R (M = 3.94, SD = 6.04); Z = -3.96, p < .001. The effect size is large, r = .883.

Results of a Shapiro-Wilk test of normality showed that SOPL (stand on preferred leg) data was not normally distributed. Thus, a Wilcoxon signed rank test was performed. Results of this test showed there was not a significant difference in SOPL between 1PP (M = 20.4 s, SD = 11.3 s) and 3PP-R (M = 21.2 s, SD = 10.7 s) conditions.

Avatar Embodiment/Agency: Recall that we used the agency sub-scale of the standardized embodiment questionnaire for this measure [28]. Results of a Shapiro-Wilk test of normality showed that avatar embodiment data was not normally distributed. Thus, a paired-samples t-test was performed. Results of this test showed there was a significant difference in embodiment ratings, with 1PP (M = 12.3, SD = 3.89) resulting in statistically significantly greater embodiment ratings over 3PP-R (M = 9.37, SD = 4.25); t = -2.55, p = .006. The effect size is medium, Cohen’s d = .614.

Usability: Results of Shapiro-Wilk tests of normality on ease and naturalness of use showed neither were normally distributed. Thus, Wilcoxon signed rank test were performed. Results of this test showed no significant difference in ease of use ratings between 1PP (M = 28.6, SD = 4.16) and 3PP-R (M = 28.6, SD = 4.20) conditions. Additionally, there was no significant difference in naturalness of use between 1PP (M = 27.7, SD = 5.40) and 3PP-R (M = 26.7, SD = 4.92) conditions.

Competence: Results of a Shapiro-Wilk test of normality showed that UPEQ Competence data was normally distributed. Thus, a paired-samples t-test was performed. Results of this test showed there was not a significant difference in UPEQ Competence ratings between 3PP-R (M = 24.4, SD = 3.27) and 1PP (M = 22.8, SD = 3.77) conditions.

Task completion time: Results of a Shapiro-Wilk test of normality showed that task completion time data was not normally distributed. Thus, a Wilcoxon signed rank test was performed. Results of this test showed no significant difference in task completion time data between 1PP (M = 96.9 s, SD = 37.4 s) and 3PP-R (M = 95.0 s, SD = 62.0 s) conditions.

Participant Preferences: The semi-structured interview shed light on what perspective people preferred and why. 18 participants stated they preferred 1PP, 15 participants stated they preferred 3PP-R, and 4 participants were unsure or thought it was situational (e.g., depends on the game). One participant was in a hurry and did not complete the interview. Participants generally stated that 1PP was more immersive than 3PP-R. 3PP-R was thought to be more disconnected from the avatar and game. However, 3PP-R was also less dizzying; this lines up with the SSQ results, which show there was a significant difference in cybersickness reported, favoring 3PP-R over 1PP. Participants generally liked the extra view of 3PP-R. They stated it was easier to see the environment, providing, as stated by P2, an "eagle eye perspective". However, this extra view could have been a detriment by removing the challenge of the maze navigation task, as participants generally stated 3PP-R "made the game very easy [P7]". Some participants, however, saw the potential of this perspective in other games to, for instance, "solve more technical puzzles [P1]".

5 EXPERIMENT 2: GAME INTERACTION

To test how well 3PP-R scales to rich and dynamic interactions typical to action games, we developed the game prototype shown in Figure 5 and on the supplemental video.

The game begins by placing the user’s avatar in a room and prompting the player to try out the locomotion and body rotation. The user then proceeds through a series of locations, each demonstrating different movements and interactions:

- Dodging projectiles by sidestepping (Figure 5A, video at 00:35). This showcases agile exaggerated locomotion and the
ability to use a death animation for the avatar if the player fails to dodge. In 1PP, such a death animation would cause nauseating viewpoint movement.

- A winding staircase, demonstrating viewpoint rotation and vertical movement (Figure 5B, video at 01:07).
- Exaggerated jumping between platforms (Figure 5C, video at 01:12).
- Grabbing and dropping objects, with target acquisition logic that counteracts the reduced interaction accuracy in 3PP compared to 1PP (e.g., [2, 30]). Objects are pulled to the user’s hand when the user presses the grab button. (Figure 5D, video at 01:23).
- Punching target dummies, demonstrating the slide-to-target exaggerated punching (Figure 5E, video at 01:46).
- Leaping to hang from a bar, swinging back and forth to gain momentum, and then leaping onto a new platform (Figure 5F, video at 02:06). This is implemented by temporarily making the avatar a physics-simulated ragdoll, connecting the ragdoll hands to the bar using ball-and-socket joints, and using shoulder joint motors controlled with the hand tracking data. Together with the exaggerated jumping and punching, this demonstrates how 3PP-R allows flexible manipulation of avatar movement without taking viewpoint control away from the user.

- Leaping onto an enemy from high above and fighting waves of enemies on an arena (video at 02:44). This demonstrates combining multiple interactions: joystick locomotion, exaggerated tracked locomotion for dodging attacks, and exaggerated punching.

We tested the prototype informally with 10 participants, one being a family member of an author, and others recruited from students and faculty of Aalto University. The participants included a professional animator / animation teacher and a professional game designer / game design teacher. Each participant played through the game, after which we asked which parts of the game were the most interesting, how the participants would improve the game, what else they would like to experience in this type of VR, and whether they experienced nausea or discomfort. Data was analyzed and discussed by two researchers, but was not coded independently.

**What was interesting?** Three primary themes were mentioned in the interviews. First, participants liked the laser dodging. This is despite the fact that some participants could not complete this section of the game or had to have assistance from the experimenters. They found it the most interesting because of that very challenge, as well as the requirement to make dynamic movements with their bodies in order to dodge. Second, participants liked the activities where it was clear that real-world movements caused the avatar to respond in the same way, e.g., the bar swinging. Finally, participants stated...
they liked the combat section of the game because it was easy, intuitive, and allowed for dynamic and varied movements. This was emphasized in particular by the animator and game designer. It was also commented that the punch exaggeration elevated the fighting to a new level, making it more spatial and dynamic.

*Suggested improvements:* Several participants reported difficulties in determining the distance between the avatar and interactive objects, such as the cubes they had to pick up, as well as bars they had to jump to and grab. This is in line with prior research on how a 3rd person perspective can make precise interactions more difficult [2, 30]. More research is needed to determine the best ways to mitigate this, e.g., by switching to 1PP for some interactions like in [32], or designing lighting and shadows in a way that gives extra depth cues. For example, a bar to jump to could cast a shadow that shows when it is directly above the player. One participant also wished to see further ahead; it may be that the 3PP-R display volume should be expanded in the direction the user and avatar are facing. Other complaints and suggestions were mostly about general usability issues, such as the laser dodging being too difficult for some players. This kind of feedback is to be expected, given that the game is a research prototype that has not yet been comprehensively tested and finetuned.

*What would the participants like to experience?* Participants generally stated they would like even more varied and dynamic movements, such as crouching, flying, swimming, or swinging around using a "ninja rope". In addition, some stated they would like a larger world to explore.

*Nausea and discomfort:* All participants stated they had no experience of nausea. Two participants mentioned having experienced nausea when they played first-person virtual reality games and said that 3PP-R worked well in that regard. The only discomfort mentioned by a participant was about the fit of the headset, as real-world jumping would sometimes slightly dislodge the headset.

6 DISCUSSION

Our experiments indicate that 3PP-R works as expected, and enables full-body interaction and agile locomotion in VR spaces larger than the real space, with little or no cybersickness. In our maze navigation study, the difference in Simulator Sickness Questionnaire scores between 3PP-R and 1PP was clear, and none of the participants reported nausea in testing the game prototype.

Naturally, we do not aim to replace all VR interaction with 3PP-R. Instead, we consider 3PP-R an additional tool for VR design, and 3PP-R could also be combined with 1PP interaction sequences. For example, an action game could use 1PP for precise object manipulation (levers, machines) and dialogue with non-player characters, and switch to 3PP-R for navigation and/or combat. User preferences based on our maze task were divided, and users also commented that the suitability of each perspective depends on the game or application.

In light of the results, the strength of 3PP-R is in games that combine interaction and agile locomotion in large spaces, e.g., in hand-to-hand arena combat. Based on our bar swinging implementation, 3PP-R is also well suited for the physically based gameplay features requested by our participants, e.g., a ninja rope, which could be highly nauseating in 1PP. Previously, interactive control interfaces for physically based avatar movement have been investigated in the computer animation literature [38]. 3PP-R appears to provide a convenient way of displaying and integrating such interaction in VR.

We also consider 3PP-R useful for video games for health and physical exercise, as turning around for real induces more body movement than only utilizing a controller, and many participants especially liked the parts of the game prototype where they could use their bodies and not just the joystick. From a health perspective, it would of course be better to be running and walking long distances physically, but this is not feasible except in warehouse-scale VR installations. 3PP-R gameplay should be able to induce substantial physical exertion even in a limited space, if using enough jumping, crouching, punching, and/or other gross-motor movements.

7 LIMITATIONS AND FUTURE WORK

We have not yet comprehensively investigated the design space of 3PP-R interaction. Our user study only manipulated viewpoint, and the experimental conditions did not include variations of other variables. In future work, we plan to conduct more experiments, investigating user preferences of the parameters in Table 1 and the effect of other modifications, such as exaggerated user rotation speed and different visualization volume sizes and shapes. For example, if one removes the cutout shader, one gains better spatial awareness; the trade-off between this and possibly greater VR sickness needs to be better understood in 3PP-R.

We also plan to test different virtual display types; with the miniature 3D model we use, it is natural to only make the display orbit around the vertical axis when the user turns. Presently, the user can’t look up at the sky of the 3rd person virtual world. However, this could possibly be implemented with a planar virtual 2D display that orbits around the user both vertically and horizontally. Furthermore, the 3PP-R setup was designed for the widely available 6 degrees of freedom head and hand tracking, but future work should investigate whether it is better to control the 3PP-R WiMi position based on head or body rotation.

In addition, given the simple nature and short duration of the maze navigation task, we did not measure presence. We expect 3PP-R to be generally similar to 3PP in that regard, eliciting less presence than 1PP [53, 57, 63]. However, 3PP-R provides more freedom of movement and more diverse embodied interaction than 3PP, with less cybersickness than 1PP. As our game prototype was praised for dynamic and varied movements, a reasonable hypothesis to test in future work is that such movement enabled by 3PP-R, as demonstrated in our game prototype, may increase presence over 3PP and joystick control and be on par with traditional 1PP.

8 CONCLUSION

We have presented 3PP-R, a simple but powerful VR display and interaction technique that allows a new range of interactive experiences with less cybersickness. 3PP-R allows interacting with natural movements (e.g., walking, turning, punching) and multiple ways of manipulating and empowering user-avatar movement mapping. Our game prototype employs techniques such as exaggerated
locomotion speed, slide-to-target punching, and joystick-based locomotion, which all reduce the need for real-world space. Thanks to the reduced visual-vestibular sensory conflict, the movement manipulation also causes less cybersickness in 3PP-R than in 1PP. Our Simulator Sickness Questionnaire (SSQ) results indicate no or very little sickness in 3PP-R maze navigation using steering locomotion, and considerably more sickness in 1PP. Also, none of our game prototype testers experienced any nausea.

The main novelty of 3PP-R is allowing the user to turn 360 degrees without losing sight of a 3rd person avatar. This is convenient with recent low-cost wireless VR headsets that allow body rotation without getting tangled with cords. Additionally, many other features come together in a novel way in our prototype. To our knowledge, no previous system has both empowered user movement in multiple ways and counteracted the associated visual-vestibular sensory conflict through limiting the 3PP view volume and using an additional 1st-person reference world.

ACKNOWLEDGEMENTS
This project was supported by Academy of Finland grant 299358. Tuukka M. Takala is a JSPS International Research Fellow.

REFERENCES


Chadwick A Wingrove, Yonica Hacidmetoglu, and Doug A Bowman. 2006. Overcoming world in miniature limitations by a scaled and scrolling WIM. In 3D User Interfaces (3DUI’06). IEEE, 11–16.