DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays

Wenge Xu
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, China
wenge.xu@xjtlu.edu.cn

Hai-Ning Liang*
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, China
haining.liang@xjtlu.edu.cn

Yuxuan Zhao
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, China
yuxuan.zhao16@student.xjtlu.edu.cn

Difeng Yu
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, China
difeng.yu@gmail.com

Diego Monteiro
Xi'an Jiaotong-Liverpool University
Suzhou, Jiangsu, China
d.monteiro@xjtlu.edu.cn

ABSTRACT

We present DMove, directional motion-based interaction for Augmented Reality (AR) Head-Mounted Displays (HMDs) that is both hands- and device-free. It uses directional walking as a way to interact with virtual objects. To use DMove, a user needs to perform directional motions such as moving one foot forward or backward. In this research, we first investigate the recognition accuracy of the motion directions of our method and the social acceptance of this type of interactions together with users' comfort rating for each direction. We then optimize its design and conduct a second study to compare DMove in task performance and user preferences (workload, motion sickness, user experience), with two approaches—Hand interaction (Meta 2-like) and Head+Hand interaction (HoloLens-like) for menu selection tasks. Based on the results of these two studies, we provide a set of guidelines for DMove and further demonstrate two applications that utilize directional motions.

CCS CONCEPTS

• **Human-centered computing** → **Mixed / augmented reality**; **Gestural input**.

*Corresponding author who can be contacted via haining.liang@xjtlu.edu.cn

CHI 2019, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5970-2/19/05...$15.00
https://doi.org/10.1145/3290605.3300674

KEYWORDS

Augmented Reality, Head-Mounted Display, Motion Direction, Menu Selection

ACM Reference Format:
Wenge Xu, Hai-Ning Liang, Yuxuan Zhao, Difeng Yu, and Diego Monteiro. 2019. DMove: Directional Motion-based Interaction for Augmented Reality Head-Mounted Displays. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland Uk.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3290605.3300674

1 INTRODUCTION

Augmented reality (AR) allows users to interact with virtual objects that are overlaid on the physical space via see-through head-mounted/worn displays (HMDs/HWDs). Ordinarily, gestural input [13, 53] is preferred to keyboard and mouse. AR HMDs have sensors that can detect head and hand movements [31, 36, 37]. What these sensors can also capture is body motion (e.g. moving the body forward/backward or left/right) by assuming that the position of the head is the position of the user and that users' head can move along with their body towards a certain direction. Unlike head- and hand-based gestures, body motion is underexplored and thus underutilized in current AR systems. Body motion can present several benefits compared to hand- and head-based motion. Hand-based motion usually requires users to keep their hands in mid-air which could result in arm fatigue during prolonged interactions [39]; it can also cause inaccurate interactions (e.g. unwanted menu item selection)—for example when users' hands accidentally go off the small tracked area of HMDs. Similarly, HMDs often cause motion sickness and, when using frequent head motions, there is the risk of increased sickness [60]. With body motion, it is possible to avoid arm fatigue and to minimize motion sickness and, as shown later in our results, still allows for high accuracy of interaction and good usability ratings. Our research explores the use of directional body motion to interact with AR HMDs

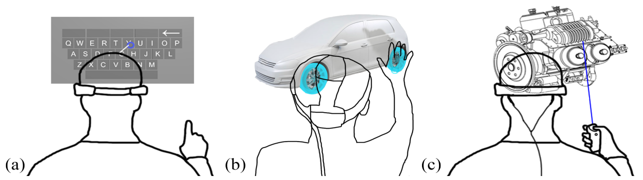


Figure 1: Interaction in three commercial HMDs, (a) HoloLens—Head+Hand-based (Hybrid) interaction (b) Meta 2—Hand-based interaction (c) Magic Leap 1—Controller-based interaction.

based on the accuracy of object selection, task completion time, and user subjective feedback on workload, motion sickness, and overall usability. Our focus in this paper is on menu item selection, but the results are applicable to other types of interaction, and interface.

At present, there are three main commercial AR HMDs—the Meta 2 [36], Magic Leap 1 [31], and HoloLens [37]. Figure 1 shows how each device supports users' interaction with the virtual environment. Meta 2 allows hand-based interaction where users need to move their hand to the menu item and confirm the selection by using a hand gesture (i.e. grab). HoloLens uses a hybrid approach for menu selection, where a ray is extended from the virtual camera position towards the viewing direction and into the virtual environment. The end of the ray is akin to a cursor and users confirm a selection by a hand/finger gesture (i.e. hand-tap)—in other words, it requires users to use their head to move the cursor and their hand for selection. This research only considers device-free approaches since they are more flexible than device-based approaches and can be used in more scenarios, environments, and types of AR devices.

In this paper, we present DMove, an approach to interact with AR HMDs that is hands-free, does not require handheld devices, and avoids the need to use head motions; instead, it uses directional body movements. In our approach, the system is trained to recognize the possible directional body motions around the user with 2 distances (Far and Close). Selection is made when the system predicts that the user has made a particular movement. Our approach only needs the sensors that already come in current AR HMDs, like Meta 2; and unlike Magic Leap, it does not require a handheld device. In the first of two studies, we explore two aspects. The first deals with the feasibility and accuracy of our recognition method, and the second is about assessing users' social acceptance of directional motion-based interactions and their perceived physical and mental comfort levels in each direction. Based on the results, we then optimize our technique and, in a second study, we compare DMove with hand-based interaction (like what is available to Meta 2 users) and Head+Hand-based interaction (akin to what users do with HoloLens). Menu selection is the chosen task because it is a common activity in AR and other types of HMDs. Based on the results of the two studies, we are able to extract a set of guidelines for interfaces that are based on directional motions. Also, we present two sample applications that can leverage DMove-type apart from menu selection.

The contributions of the paper include: (1) a motion direction recognition method that requires no additional handheld devices nor sensors for current AR HMDs; (2) an optimized directional motion-based interface (DMove); (3) an evaluation of 3 menu selection methods for AR HMDs; (4) a set of guidelines for applications that use directional motion-based interactions; and (5) two applications external to menu selection and that use DMove as their interaction interface.

2 RELATED WORK

Device-free Interaction in AR HMDs

Mine [38] pointed out that interacting with virtual objects requires (1) a mechanism for the identification of the objects to be selected, and (2) some signal or command to indicate their selection. We next describe two commonly used device-free interactions for AR HMDs.

Hand-based Interaction. Hand-based interaction is one of the most commonly used selection methods in AR HMDs [34] because it is assumed to be natural and practical. To perform a selection of a near object [38], users first need to choose the virtual object to be selected by hovering the hand over it and then selecting it by performing a gesture—e.g. in Meta 2 [36] users select the item by making a grab gesture. To select an item that is placed further away from the user, Mine [38] suggests that users can utilize their finger to point at the object followed by a selection gesture. Studies have looked at the finger pointing [4, 33], but these techniques require an additional external sensor like Kinect that is placed at a distance to detect and classify the gestures.

In general, hand-based interactions that require users to hold their hands in mid-air are uncomfortable and can be tiring, particularly for AR/VR devices [51]. This is because users are forced to keep their hands within the small area tracked by the sensors. Inaccuracies can often occur when the hands go off the area. In addition to issues with the recognition algorithm and other technical limitations [55], mid-air hand interactions are also sensitive to users' physical abilities which can lead to unpredictable performance.

Head-Pointing. Together with hand-based techniques, head-based interaction has been actively studied in the virtual reality (VR) HMDs [7, 10]. It has been widely adopted as a standard way of pointing at virtual objects without using hands or hand-held pointing devices [29]. Instead, it relies on the HMDs' built-in IMU sensors. Recent studies further have explored head-based techniques in both VR [3] and AR [29]. Like techniques based on eye-gaze, using the head may lead users to suffer the 'Midas Touch' [27] problem of unintentional selection because head-pointing has this same problem when confirmation of a selection is needed. Researchers have investigated solutions to this problem such as using dwell time [27, 41, 49, 54], adopting gaze gestures [9, 12, 25, 26], applying a second modality such as controllers [29], but these solutions are at times not ideal. For example, having a dwell time can slow performance; gaze requires additional expensive trackers but still suffers from accuracy issues; and not every AR HMD can track a handheld device, furthermore forcing users to hold a device prevents their hands from being used to manipulate the virtual objects in these systems.

One solution used in commercial HMDs is combining both head and hand-which is referred to as hybrid interaction, which relies on the use of the head to move the cursor to a target and hand gestures to confirm the selection, like it is done with HoloLens [37]. However, this approach still suffers from the limitations of hand-based interaction.

Body Motion-based Interaction

Foot-based Interaction. Alexander et al. [2] suggest that foot-based interactions can be grouped into two categories based on how foot actions are mapped to system commands. *Discrete* foot gesture [11, 47, 59] are those that are mapped to specific tasks (e.g. locking and unlocking a mobile phone). *Continuous* gestures [19, 24, 40, 43, 44, 46] are those that are mapped to tasks with a spatial component (e.g. moving in one direction in a space). Although it can add an extra dimension to users' interaction, in general the proposed techniques using users' feet require additional external sensors. This constraint limits users to fixed environments and within the space tracked by the sensors. Because AR HMDs are meant to allow freedom of movement, the need to have external sensors is not desirable. Our approach avoids this constraint and relies solely on the sensors that already come with current commercial AR HMDs.

Full Body-based Interaction. Body motion direction-based interactions have several advantages. As our results show, they can be accurately predicted by a system that requires minimal training. They avoid the pitfalls of hand- and head-based interaction. Body motion tends to be natural and does not force users to be in uncomfortable, unnatural positions for long periods (like hand interactions which users must hold their hands in mid-air). Also, as our results show, they do not increase motion sickness despite the need for users to make body movements.

Given their potential benefits, but without the limitations of other types of gestures, we want to explore the use of the

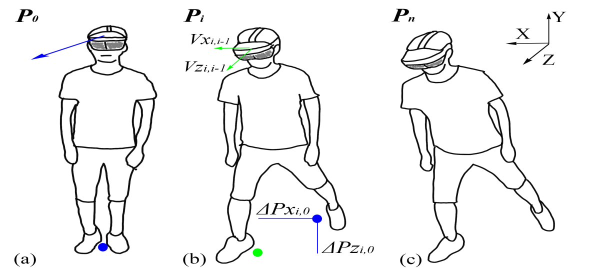


Figure 2: An example of a movement. (a) Starting State—A user is ready to move toward the North-East direction. The blue dot is the starting position tracked by the system. (b) Prediction State—The state used to predict the moving direction where the user has nearly finished the movement. The green dot is the end position tracked by the system. The system calculates $v_{x_{j,j-1}}$, $v_{zj,j-1}$, $\Delta P_{x_{i,0}}$, $\Delta P_{z_{i,0}}$ and then sends the results to the algorithm. (c) End State—A movement is finished.

motion-based interactions for current consumer AR HMDs. We also want this type of interaction to be hands- and device-free. As this research shows, our technique DMove is as fast as other methods for menu item selection and also brings a subjectively better user experience.

Besides, full body motion-based interactions can be applied to other domains (e.g. gameplay [14, 42]). Further, this type of interaction can encourage physical activity in offices and homes and as such can bring health benefits to their users—e.g. just ten minutes of physical activity can help users gain cognitive and physical benefits [28]. Besides work-related applications, body motion can be used for gaming interfaces. For instance, an exergame leveraging body motion as input has the potential to be utilized to encourage physical activity, so that for example elderly users or children can do exercises in a fun way regularly at home to develop their physical strength [16, 52]. At the end of the paper, we present a sample of exergame that uses motion-based interactions.

3 DMOVE

In this section, we discuss the DMove's motion recognition method and the interface for our Study One.

Motion Recognition Method

We use machine learning to classify the user's motion direction. Instead of classifying it through movement patterns (i.e. changes in the sensors' acceleration values in three dimensions), we identify the gesture (i.e. posture at the end of the movement). This is because the former approach will not always work because some HMDs, like Meta 2, do not allow access to their acceleration data; we want to make this method suitable for all AR HMDs.

Next, we describe our method in detail. In three dimensions, a spatial position is defined as $P = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ (see Figure 2). A spatial path $\prod$ describes the spatial progression of movement. It is an ordered list of measured spatial positions: $\prod = (P_0, ..., P_i, ..., P_n)$; where P_0 is the starting position, P_i is the position we predict the performed movement, P_n is the position where the user finishes the motion (see Figure 2). The values used in our analysis process are described in the following formulas:

$$\Delta P_{x_{i,0}} = P_{x_i} - P_{x_0} \tag{1}$$

Where $\Delta P_{x_{i,0}}$ is the distance users moved/traveled from the starting position (P_0) to the position to be predicted (P_i). This formula also applies to $\Delta P_{z_{i,0}}$.

$$\Delta v_{x_{j,j-1}} = \frac{\Delta P_{x_{j,j-1}}}{\Delta t_{j,j-1}} \tag{2}$$

Where $\Delta v_{x_{j,j-1}}$ is the current speed of the head along the X-axis. $\Delta P_{x_{j,j-1}}$ and $\Delta t_{j,j-1}$ are the distance and time differences between this frame and the respective last frame. This formula also applies to $\Delta v_{z_{j,j-1}}$.

$$m = \frac{\Delta P_{x_{i,0}}}{\Delta P_{z_{i,0}}} \tag{3}$$

Where m is the slope of the line from P_0 to P_i in X-axis and Z-axis.

Classification. Tested features are $\Delta P_{x_{i,0}}$, $\Delta P_{y_{i,0}}$, $\Delta P_{z_{i,0}}$, distance traveled between P_0 and P_i, slope m. Only $\Delta P_{x_{i,0}}$ and $\Delta P_{z_{i,0}}$ are included in our dataset since the features analysis using Weka [18, 56] has shown that they are the top 2 features and all predictions are based on them. We apply the Random Forest classifier provided by Weka for predicting the motion directions. Figure 3 shows the algorithm flowchart.

Interface and GUI

We proposed two interfaces that are based on eight directions—East (E), North-East (NE), North (N), North-West (NW), West (W), South-West (SW), South (S), and South-East (SE). Figure 4 shows the two designs. The first design is 8-block DMove which each direction has one distance level—No Limit (we suggest at least 20 cm away from the starting position to improve the accuracy); the other is 16-block DMove which each direction has two distance levels—Close (we suggest 30 cm away from the starting position) and Far (we suggest 60 cm away from the starting position). We wanted to use two levels of the distance (Far and Close) around the user because, with two levels, the technique can have more interface items, but this may also affect the prediction accuracy of distinguishing between the two levels. To guide users visually, both interfaces are displayed in front of their view

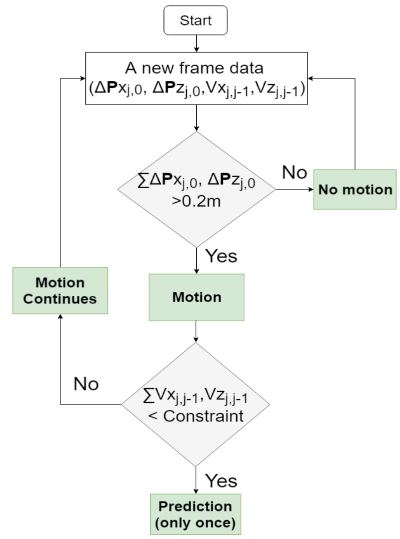


Figure 3: Algorithm flowchart for predicting the motion direction; we set the constraint to 0.1 m/s since it works well according to our test trials.

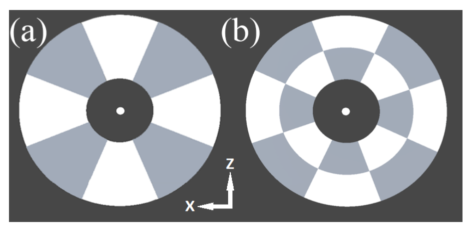


Figure 4: (a) 8-block and (b) 16-block DMove interface.

like a GUI where the tiny white point in Figure 4 represents the head position.

4 STUDY ONE

In this study, we focused on the accuracy of our motion direction recognition technique. We also investigated the social acceptance of the motions (i.e. in front of whom users would accept to perform these motions and where) and comfort levels (mental and physical) of doing such motions.

Participants and Apparatus

12 participants (4 female) aged between 17 and 28 were recruited from a local university campus to participate in the study. They all had normal or corrected-to-normal vision. The study was conducted using a Meta 2 AR HMD [36] connected to a standard computer with an i7 CPU, 16 GB RAM and an Nvidia GeForce GTX 1080Ti GPU. We implemented the system in Unity3D. All experiments were conducted in a lab where users cannot be seen from outside.

Q1) On a scale from 1 to 6, what was your overall impression/emotion during the task?

1☐ 2☐ 3☐ 4☐ 5☐ 6☐

I hated it, it felt terribly awkward — I enjoyed it, it felt comfortable

Q2) Imagine that this motion direction gestures can be used to control a menu or play a dance game. Now, **in front of whom** do you think you would **feel comfortable** using such gestures? Select **one or more** items from the list below.

☐ I would **not** feel comfortable using them even when alone
or
☐ when alone ☐ in front of my partner ☐ in front of friends
☐ in front of family ☐ in front of colleagues ☐ in front of strangers

Q3) Now, **in which locations** do you think you would **feel comfortable** using such gestures? Select **one or more** items from the list below.

☐ I would **not** feel comfortable using them no matter where I am
or
☐ at home ☐ on the sidewalk ☐ in a pub, café, or restaurant
☐ in a shop ☐ in a museum ☐ as a passenger on a bus or train
☐ at my workplace

Figure 5: Sample of social acceptance questions (adapted from [1]).

Design and Evaluation Metrics

The experiment employed a one-way within-subjects design where the independent variable was interface—16-block and 8-block. We were interested in two variables, (1) *Target Direction*—E, NE, N, NW, W, SW, S, SE; and (2) *Target Distance*—Close, Far, and No Limit. Participants were asked to do a training data collection session first for both interfaces and then do the testing sessions. The order of the interface was counterbalanced.

The evaluation metrics were listed below:

- *Accuracy.* Accuracy was measured based on reproducibility [17] and how stable and scalable the system was against the data collected from a different session. An error was recorded when the classifier failed to predict the correct movement direction.
- *Physical and Mental Comfort.* It quantified how the users' comfort levels (both physical and mental) varied across each Target Direction × Target Distance combination. We used 5-point Likert questions to collect the data.
  There are measurement tools for fatigue and comfort. And one of the most common one s is the 6-20 Borg scale [6]. However, we did not choose to use this scale because it seemed more suitable for measuring physical comfort alone. The questionnaire we used measured both physical and mental comfort of the participants during their interaction with the AR HMD.
- *Social Acceptability.* We adopted the questionnaire from [1] (see Figure 5) to assess in which places and in front of whom users were comfortable doing the motions.

Task and Procedure

The experiment began with the data collection session for each interface where the order of the interface was counterbalanced. The system would ask participants to perform each directional movement five times starting from N followed by the other directions in a clockwise order till the last direction, i.e. NE→E→SE→S→SW→W→NW. For the 16-block DMove, the system would ask participants to do the Target Direction × Close first then Far. For the 8-block DMove, they only needed to do the No Limit movement for each direction. They were asked to let the head follow their body movement in a natural way to help them keep their balance and their head steady. In between conditions, participants were requested to fill out the Physical/Mental Comfort questionnaire.

After the data collection session, they did the testing session. The order of interfaces was the same as the data collection session for each participant. However, unlike the testing session, which had a fixed order for the direction, in this phase, the system randomized the directions. This was done to better assess the accuracy of the system and to avoid participants' muscle memory. Similar to the data collection session, participants had to reach each direction five times.

At the end of the experiment, participants completed the social acceptability questionnaire. The whole experiment lasts around 30 minutes for each participant.

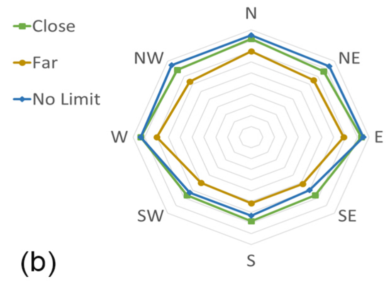
Results

Accuracy. We used 2880 instances collected from the training session to train the model and used another 2880 instances from the testing session to test it. The accuracy, precision, recall, F-Measure for 8-block DMove were all 100% while for 16-block were 98.06%, 98.2%, 98.1%, 98.0%, respectively. As can be observed from the red blocks of the confusion matrix in Figure 6a, most of the wrong predictions were in South Close where our recognition method predicted South Close as South Far.

Subjective Feedback. The collected data were analyzed using a two-way repeated measures ANOVA with two factors (1) Target Location and (2) Target Distance. Bonferroni corrections were used for pairwise comparisons. For violations of sphericity, we used a Greenhouse-Geisser adjustment for degrees of freedom.

Physical Comfort. Figure 6b shows the Physical Comfort ratings of each direction for Target Distance. An ANOVA showed significant effects of Target Direction ($F_{3.029,16.737} = 11.130, p < .001$) and Target Distance ($F_{1.860,20.458} = 13.899, p < .001$) on Physical Comfort.However, no significant interaction effect of Target Direction × Target Distance ($F_{14,154} = 1.076, p = .383$) was found. For Target Direction, post-hoc pairwise comparisons revealed significant differences between N-SW, E-SE, E-S, E-SW, SE-W, SW-W, SW-NW (all $p < .05$). It also yielded a close significant difference between N-SE ($p = .073$), N-S ($p = .076$), SE-NW ($p = .053$),

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
	96.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	A = N-Close
	0.0	94.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.6	0.0	0.0	0.0	0.0	0.0	0.0	B = NE-Close
	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	C = E-Close
	0.0	0.0	0.0	98.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	0.0	0.0	0.0	0.0	D = SE-Close
	0.0	0.0	0.0	0.0	83.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.1	0.0	0.0	0.0	E = S-Close
	0.0	0.0	0.0	0.0	0.0	97.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.2	0.0	0.0	F = SW-Close
	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	G =W-Close
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.7	H = NW-Close
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	I = N-Far
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	J = NE-Far
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	K = E-Far
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	L = SE-Far
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	M = S-Far
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	N = SW-Far
(a)	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.4	0.0	O = W-Far
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	P = NW-Far



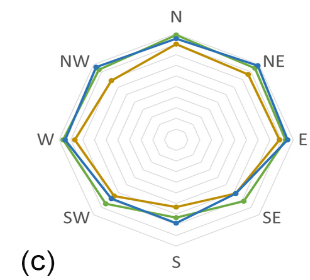


Figure 6: 16-block DMove Confusion matrix (a). Comfort ratings for each direction for Physical (b) and Mental (c).

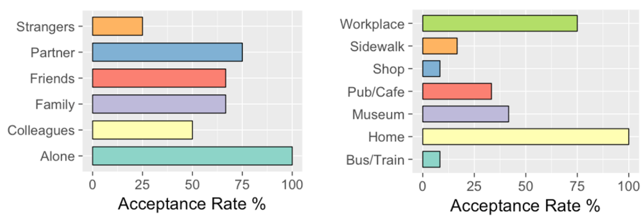


Figure 7: Acceptance rates for different audiences (a; left), and locations (b; right).

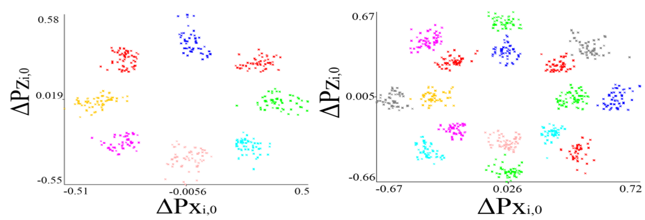


Figure 8: 8-block (a; left) and 16-block (b; right) DMove's plot image of $\Delta P_{x_{i,0}}$, $\Delta P_{z_{i,0}}$, where each color represents a movement direction.

and S-W ($p = .063$). For Target Distance, pairwise comparisons revealed a significant difference between Close and Far ($p = .001$), No Limit and Far ($p = .005$), but not between Close and No Limit ($p = 1.000$).

Mental Comfort. Figure 6c shows the Mental Comfort ratings of each direction for Target Distance. An ANOVA yield a significant effect of Target Direction ($F_{2.420,26.619} = 17.492, p < .001$) and Target Distance ($F_{2,22=8.305}, p < .05$) on Mental Comfort. However, there was no significant interaction effect of Target Direction × Target Distance ($F_{4.032,44.355} = 1.868$, $p = .132$). For Target Direction, post-hoc pairwise comparisons revealed significant differences between N-SE, N-S, NE-SE, NE-S, NE-SW, E-SE, E-S, E-SW, SE-W, SE-NW, S-W, S-NW, SW-NW (all $p < .05$). For Target Distance, pairwise comparisons revealed a significant difference between Close and Far, No Limit and Far (both $p < .05$) but there was no significant difference between Close and No Limit ($p = 1.000$).

Social Acceptability. Participants' overall feelings during the task were rated 4.5 out of 6 (s.e. = .195). We calculated the acceptance rate for each given audience and location using the percentage of participants who selected each audience/location in their answers (see Figure 7). A Cochran's Q test showed a significant difference between audiences ($\chi^2(5) = 20.606, p < .001$). Post-hoc McNemar tests (Bonferroni: α-levels from 0.05 to 0.004) showed that the acceptance rates for strangers were significantly lower than if participants were alone ($p < .004$). Also, participants' responses suggested that the location would influence their willingness to use directional motions. A Cochran's Q test showed a significant difference between locations ($\chi^2(6) = 39.368, p < .001$). Post-hoc McNemar tests (Bonferroni: α-levels from 0.05 to 0.004) showed that the acceptance rates for using DMove at home was significantly higher than at a shop or other public places, and on sidewalks (all $p < .004$).

Discussion

Direction Motion-based Interface. Our method showed very good accuracy for identifying the users' movement direction in both 8- and 16-block DMove interfaces. The reason was that the attributes used in our dataset clearly distinguished the movement directions (see Figure 8). Participants' subjective feedback indicated that motions toward the South direction lead to both physical and mental discomfort. During the experiment, we also observed that each participant had his or her own predisposed way of making directional movements due to their physical attributes—e.g. taller users were able to take a longer step than the shorter users. As such, we believe that using a user's own motion data will likely increase prediction performance because it will consider the physical characteristics of each participant.

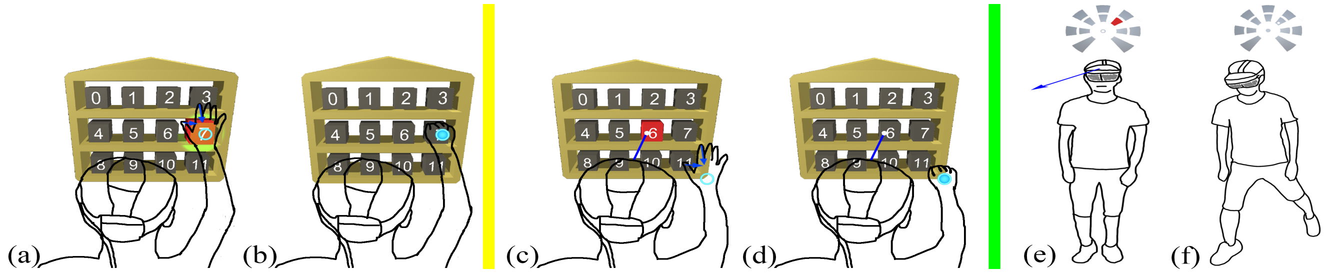


Figure 9: Using Hand, Hybrid, and DMove to select an item from the menu. (1) Hand (a) A user needs to move the hand to the target and hover it, (b) and then performs a close palm gesture to select it. (2) Hybrid (c) A user needs to rotate the head to move the cursor to the target, (d) and then performs a palm closing gesture to select it. (3) DMove (e) A user needs to go the NE direction, (f) a selection is made when the user (nearly) completes the action.

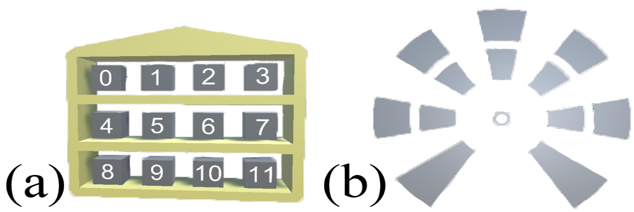


Figure 10: (a) Hand/Hybrid—Meta 2 Workspace-like menu interface, and (b) final-DMove interface—optimized based on the 16-block layout with S removed and had one single larger area for SE and SW directions (only 1 level due to users' discomfort with two levels).

Social Acceptability. According to the results of the social acceptability questionnaire, most participants were quite positive towards a DMove-based interface; only one participant gave a low rating of 3. They were willing to do directional motions alone or in front of familiar people (see Figure 7a). They preferred private spaces (such as their home and workplace) rather than public areas (see Figure 7b). Based on this feedback, we suggest that a DMove-type of interface should be used in in-door scenarios (i.e. home or workplace) and in front of people familiar to the user.

Optimization. Based on the performance and subjective feedback, we decided to work further with the 16-block interface and optimize it. Since users have difficulty moving towards the S direction, we decided to make some adjustments to S and also SE and SW directions. We removed S and combined the 2-levels SE and SW directions into one single direction each. In this way, users could easily move towards these two (now much larger) directions. After these changes, the DMove interface had 12 items (Figure 10b).

5 STUDY TWO

In the second study, we explored the use of DMove for menu selection, a very common activity in AR HMDs. We compared the performance, suitability, and usability of DMove with two device-free interaction methods, Hand-based and Hybrid (Head+Hand), for menu selection because they represent two of the most common, and available ways for selecting menu items in current AR devices. Similar to Study One, we also measured workload, motion sickness, and user experience of the three methods. We only considered device-free approaches because they are applicable to a wider range of scenarios, and types of HMDs.

Evaluated Conditions

We evaluated the following 3 Selection Methods for menu selection:

- *Hand-based interaction (or simply Hand).* This was similar to what Meta 2 would provide. To select an item in a menu, a user had to move the cursor controlled by one hand in mid-air to hover it on the item and then make a palm closing gesture to confirm its selection. Figure 9(1) shows this scenario. Visual feedback, in the form of extra green light and enlarged item, was provided to indicate whether the hand was correctly positioned on the item. A sound would be played to confirm the selection. We modified the code from one of the sample demos provided by Meta Company, the developers of the Meta 2.
- *Hybrid-based interaction (or simply Hybrid).* This was analogous to how menu selection was done in HoloLens, where a user had to move the head to control a cursor and position it on an item—selection was confirmed by a hand gesture. The HMD would track the head motion casting a ray to the virtual environment. The end of the ray was akin to a cursor, which served as visual feedback. Hand detection cursor was provided to

inform the user of the cursor's state. A sound would be played when a selection was made. Figure 9(2) shows an example of this approach.

- *Directional Motion-based interaction (DMove).* In this condition, a user had to move their body with one foot towards a direction location that represented a menu item. For any motion performed, the classifier would return the direction and block. A cursor presenting the user's position was provided on the HMD as visual feedback and a sound would be played if a selection was made. Figure 9(3) shows an example of how a user would select the NE item.

We designed the menu items (see Figure 10) based on official design guidelines [35], which suggested that they should be located at around 0.5m away from the user. However, regardless of this, the users could still adjust the position between them and the menu items to a comfortable distance before the start of the experiment. We used grid menu layout for Hand and Hybrid interaction because both HoloLens and Meta 2 have applications that rely this type of layout. For example, the developers of Meta 2 provided guidelines and an official application using a grid layout—we followed the guidelines and adapted the application for this experiment. We did not use the grid layout for DMove because it did not represent a natural mapping for around body interactions. Our choice of radial layout was based on feedback from a pilot study and also from previous research [21, 32].

Participants and Apparatus

18 participants (6 female) aged between 17 and 28 were recruited from the same local university campus as in Study One. They all had normal or corrected-to-normal vision and were right-handed. To avoid biases, none of these participants did Study One. This experiment used the same apparatus and lab location as Study One.

Experiment Design, Task, and Procedure

The experiment followed a 3×2 within-subjects design with two factors: Selection Method (Hybrid, Hand, and DMove) and Menu Size (Large—same size as Meta 2 Workspace, and Small—80% of the Large). The combinations of Selection Method × Menu Size were counterbalanced. The whole experiment lasted about one hour for each participant. Before the trials started, the participants were asked to complete a pre-experiment questionnaire to gather demographic information and were informed of the purpose of the study. Since Study One suggested that using the user's dataset could help improve recognition accuracy, we collected data from each user before the first testing session to train our system. This data collection session was conducted in the same way as in Study One but with fewer directions and took just around 2-4 minutes. To balance the conditions, participants were also given up to 5 minutes of training with both Hand and Hybrid interactions. When participants felt rested and ready, they would proceed to the testing session.

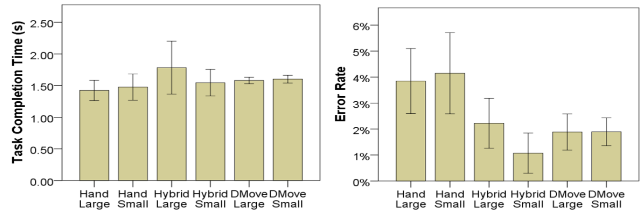


Figure 11: Mean task completion time (a; left) and error rate for the six layouts (b; right). Error bars indicate ± 2 standard errors.

In each session, each block (representing a menu item) would randomly appear once, one by one, for a total of five times. After each session participants completed three questionnaires: NASA-TLX [20], user experience questionnaire (UEQ) [30], and motion sickness assessment (MSAQ) [15]. We instructed participants to maintain their head steady and in a comfortable position whenever possible. In the end, we asked them to provide comments on each of the interfaces. The experiment returned 3 (Selection Method) × 2 (Menu Size) × 12 (blocks) × 5 (times) × 18 (participants) = 6480 trials.

Results

We analyzed the data using a two-way repeated measures ANOVA with two independent variables, Selection Method (Hand, Hybrid, DMove) and Menu Size (Large and Small). Bonferroni correction was used for pairwise comparisons, and Greenhouse-Geisser adjustment was used for degrees of freedom for violations of sphericity.

Task Performance. Figure 11 presents the task completion time and error rate among the six layouts. For task completion time, the ANOVA test yielded no significant effect of Selection Method ($F_{1.197,20.341} = 2.555, p = .121$), Menu Size ($F_{1,17} = 1.108, p = .307$), and Selection Method × Menu Size ($F_{1.219,20.715} = 1.177, p = .303$), which showed that the completion time for each Selection Method was equal. For error rate, there was a significant main effect of Selection Method ($F_{1.506,25.610} = 14.138, p < .001$), but no significant main effect of Menu Size ($F_{1,17} = .524, p = .479$) and no significant interaction effect of Selection Method × Menu Size ($F_{1.940,32.980} = 2.069, p = .144$). Post-hoc pairwise comparison revealed a significant difference between Hand and Hybrid, Hand and DMove (both $p < .05$); this meant that hand had higher error rates than Hybrid and DMove. There was no significant difference between Hybrid and DMove.

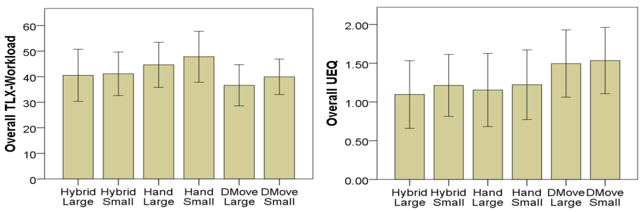


Figure 12: Overall NASA-TLX workload (a; left) and overall UEQ scores among all 6 layouts (b; right). Error bars indicate ± 2 standard errors.

NASA-TLX Workload. For overall workload, DMove Large was rated the best (M = 36.63, SD = 17.07) and Hand Small (M = 47.80, SD = 21.13) was rated the worst. ANOVA tests yielded a significant effect of Selection Method ($F_{1.514,25.732} = 4.676, p < .05$), but not of Menu Size ($F_{1,17} = 2.806, p = .112$) and Selection Method × Menu Size ($F_{2,34} = .211, p = .811$). Post-hoc pairwise comparisons revealed a significant difference between Hybrid and Hand, DMove and Hand (both $p < .05$; see Figure 12a).

Regarding NASA-TLX workload subscales, ANOVA tests yielded a close significant effect of Selection Method ($F_{2,34} = 2.947, p = .066$) on Mental; a close significant effect of Selection Method ($F_{2,34} = 2.927, p = .067$) on Temporal; a close significant effect of Selection Method ($F_{1.544,26.240} = 3.533, p = .054$) on Frustration; and a close significant effect of Selection Method ($F_{2,34} = 3.094, p = .058$) on Effort. No other significant or close significant effects were found.

User Experience. The score for UEQ was adjusted between -3 (very bad) to 3 (excellent). Figure 12b shows the overall UEQ score among the six layouts. ANOVA tests yielded a significant effect of Selection Method ($F_{2,34} = 6.371, p < .01$), but not of Menu Size ($F_{1,17} = 2.498, p = .132$). No significant interaction effect was found on Selection Method × Menu Size ($F_{1.350,22.956} = .202, p = .730$). Post-hoc pairwise comparisons showed a significant difference between Hybrid and DMove as well as Hand and DMove (both $p < .05$).

Regarding the UEQ subscales, ANOVA tests yielded a significant main effect of Selection Method ($F_{2,34} = 6.167, p < .01$) on attractiveness. The pairwise comparison indicated DMove was more attractive than both Hand and Hybrid (both $p < .05$). There was a significant effect of Menu size ($F_{1,17} = 6.115, p < .05$) on stimulation. Post-hoc pairwise comparison showed Small Menu bought more stimulation from users than Large Menu ($p < .05$). No other significant effects were found. DMove outperformed Hand, Hybrid across the UEQ subscales (see Figure 13).

Motion Sickness. For the overall sickness score, DMove Small was rated the worst (M = 19.29%, SD = 12.55%) and Hand Small was rated the best (M = 16.59%, SD = 8.73%). ANOVA tests yielded no significant effect of Selection Method ($F_{1.207, 20.521} = 2.860, p = .100$), Menu Size ($F_{1,17} = 1.569, p = .227$), and Selection Method × Menu Size ($F_{1.390,23.626} = 1.224, p = .297$) on overall motion sickness. Regarding MSAQ subscales (gastrointestinal, central, peripheral, sopite-related), ANOVA tests yielded a significant effect of Selection Method ($F_{2,34} = 4.265, p < .05$) on peripheral and a close significant main effect of Selection Method ($F_{1.149,19.532} = 4.022, p = .054$) on central. No other significant effects were found. Post-hoc pairwise comparisons showed no significant effect between any Selection Method on peripheral.

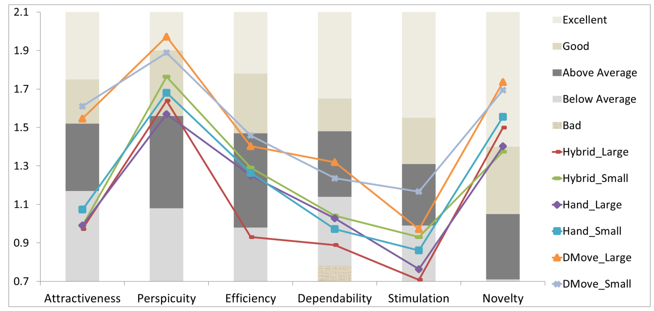


Figure 13: User Experience Questionnaire ratings for all 6 layouts with respect to benchmarks.

6 DISCUSSION

In this section, we discuss the reasons why DMove is a strong candidate interface for menu selection based on users' performance and experience for the current AR HMDs.

Task Performance

The results indicated that Hand, Hybrid, DMove have equal selection time, while Hybrid and DMove had lower error rates than Hand. We observed that the high error rate in Hand was due to wrong selection of the item that was next to the intended targets. Although visual feedback was provided (by expanding the size and adding additional highlight color) for the item that the users' hands were currently hovering on, the system's detection time for whether their hands were on the virtual item was slow (1-2 seconds). Faced with this, users chose to trust their spatial knowledge and performed the selection gesture which was often incorrect, and this led to higher error rates. This was not the case for Hybrid which the users' hands were only used to perform a gesture to confirm a selection. Interestingly, we found that Menu Size had no effect on task performance. This might have been because the difference between Large and Small was not big enough to cause a significance. Based on performance alone, we suggest Hybrid and DMove should be considered before Hand for current AR HMDs.

User Preference

NASA-TLX Workload. Regarding the overall workload, Hand was worse than Hybrid and DMove. One reason why participant felt that the overall workload was higher for Hand was that to use it well they had to focus very carefully to gauge where the items were located and the location of the virtual cursor. This process was tiresome. Although there was no difference in physical workload among three methods, participants had arm fatigue in both Hand and Hybrid—several them said it was too difficult and tiring to hold their hand for long periods or to perform the hand gesture repeatedly to make a selection. In contrast, for DMove there was no need to exercise the visual focus required in Hand because they could rely on their spatial awareness of the location of the items around them to make a quick motion for their selection. So, users should avoid using the Hand approach if they consider workload to be a crucial factor.

Motion Sickness. Our results indicated that performing directional movements in DMove did not result in a higher motion sickness than selecting menu items via Hand and Hybrid. Thus, in terms of motion sickness, we believe DMove was as comfortable as Hand and Hybrid.

User Experience. ANOVA tests showed that DMove provided a better user experience than Hand and Hybrid. As mentioned earlier, we considered Hand and Hybrid because they were used in current the AR HMDs and presumably were thought to be usable. Our results showed that only DMove was rated above average to excellent while Hand or Hybrid was rated much worse. Although our data samples were not sufficient enough to compare with the benchmarks [45], they still provided a sense of how much more usable DMove would likely be when compared to the other two interfaces. In summary, using DMove results in better user experience than Hand and Hybrid, and if users regard usability and user experience as the most important factors, DMove is the recommended choice.

User Comments

According to Bowman et al. [8], natural interactions (like Hand in our study) provide little additional productivity but actually can make the task more complicated and unnecessarily cumbersome. Hand interaction not only caused some physical discomfort and arm pain (P7: "*my arms are sores after a while*") but participants did not like it because of the lack of tactile feedback (P10: "*It feels empty when I use my hand to select the virtual objects, because I don't sense when the action is finished*"). Physical issues are not easy to solve—the only way is to ask users to rest. The tactile feedback issue could be solved by using a haptic glove. However, it is expensive. In the case of Hybrid interaction, participants seem generally happy with its task performance, but it seems to be bored and may also cause issues like arm muscle tiredness and pain (P3: "*In the end, I felt a bit sleepy and my arms get tired fast*"). On the other hand, participants have found DMove interesting and very easy to use. Participants suggested that we develop an exergame (like [42]) based on DMove, as eloquently put by P9: "*[DMove] is fun, I would recommend using it as an exergame, it's good for health*".

Design Guidelines for DMove Interactions

Guideline 1: Cater to Individual Differences. Based on our findings from Study One, DMove should use an individual's dataset to maintain (100% or close to 100%) accuracy to take into account each user's height, weight, movement speed, and step distance. To account for these factors and to prevent poor accuracy, DMove for general users should be calibrated according to individual physical features and abilities. Besides, we predict a motion just right before a user finishes it by comparing the head movement speed with a pre-set constraint, which should also be tuned to suit the individuals. As our second study show, training the system is easy and fast and needs to be done only once.

Guideline 2: Flexibility, Efficiency of Use, Customizability. The comfort ratings from Study One suggests that the Close level is much easier to reach, and it does not cause discomfort, while directions that users can see—N, NE, NW, E, W are much easier to perform. As such, we suggest putting frequently used items/functions in Close directions and avoid putting them at the directions that users cannot see easily to increase efficiency and usability.

Guideline 3: Not in Front of Strangers and Public Venues. Based on the social acceptance results from the Study One, we recommend using DMove for indoor scenarios such as at home/work environment (or outdoor but when there is nobody around). In addition, we suggest that an interface based on DMove should be used in front of the people users are familiar with instead of strangers.

Guideline 4: Provide Feedback and Keep Consistency With Other Interfaces. Results from Study Two point out two advantages of DMove over Hand and Hybrid. On the one hand, DMove provides users actual tactile feedback when they select an item/function because when placing the foot on the ground they will receive immediate and clear feedback. On the other hand, DMove is an interface that can be considered eyes-free because users can use their spatial awareness and memory to remember where the items are around them. Although it can be eyes-free, we suggest that the menu should always appear as a simple non-obtrusive visual interface on the HMD on-demand, similar to a context menu, whenever users want to use it and so that they do not have to memorize the items of the menu. Similar to what we have done in

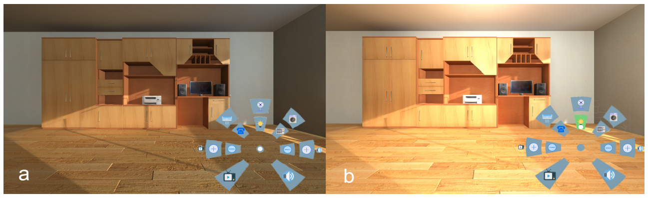


Figure 14: An example of a smart environment remote control using an AR HMD; a user realizes the environment is dark (a;left) so he/she uses remote control to switch the lights on (b;right).

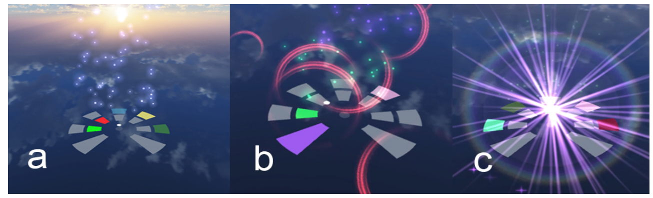


Figure 15: An example of a dance game; the DMove interface (a;left), the particle effect when a correct movement is made (b;mid), (c;right) to help users engage with the game.

this research, we suggest that the interface shows the user's movement location—e.g. a simple visual cue like a dot can be used to indicate to which direction they are moving. Visual and/or audio feedback can be included to tell them that a selection has been successfully made.

7 SAMPLE APPLICATIONS

In this section, we present two applications where DMove can be used for not only AR but also possible for VR/MR systems.

Remote Control of an Environment

We developed a prototype application (Figure 14) to remotely control electrical appliances and devices in an environment (i.e. home/workplace). There are existing methods for controlling home appliances via voice or a smartphone; however, such methods have limitations—they either are affected by ambient noise [23] or require users to have access to an additional device. DMove does not have any of these limitations. Users can use it to control smart IoT-linked devices such as a TV, lights, air condition, with a DMove-type interface. For instance, when using an AR HMD, a user realizes that the light in the room is too dark (Figure 14a), then he/she can take a small step forward, to turn the light on (Figure 14b). Further, the user is not limited to turning devices on/off only but can also to interact with a smart TV, for instance, to switch channels by taking a small step leftward and staying at "-" icon to continuously change the channels until the TV shows the desired one. If the items are not in the current interface, users can add a new item and customize its function.

Dance Exergame

Our second prototype application is a dance exergame, which can be accessed and played via a DMove-type interface. Such a game can be helpful for users of all ages to entrain themselves while doing exercise and in the process to improve their health [5, 22, 48, 50, 57, 58]. The game starts with the system randomly activating some blocks (see Figure 15). To deactivate a block successfully, the user needs to perform the corresponding directional motion within a time period, which can be adjusted based on difficulty levels. If the user fails to move and tap on the blocks before the time limit expires, the user cannot get points, which are needed to move to other levels. To avoid motion sickness, we allow users to set a time limit per round of gameplay (e.g. about 3-5 minutes akin to the length of a typical song). To make the game suitable for the elderly, one can follow the recommended guidelines provided (e.g. in [14]). In addition, the game can be multiplayer based and be played with friends via an online platform, so it could potentially bring in a social component into the gameplay. Overall, our second prototype is a dance exergame that can be played in an office or home environment with an AR HMD and potentially for a VR system as well.

8 LIMITATIONS AND FUTURE WORK

Although DMove does not cause arm and neck fatigue, repeated use in a long period may cause some degree of tiredness in the user's leg or body. On the other hand, the AR HMDs are commonly used by users in standing position. Also, as indicated earlier, standing and moving one's body is often encouraged in today's sedentary society—e.g. standing rather than only sitting while typing. As such, DMove may offer extra benefits in the form of physical activity.

As stated earlier, we have selected the grid menu for Hand and Hybrid interactions based on example applications used in two current AR HMDs. It can be argued that their layout or the items can be made smaller so that they can fit better in the common small field of view of AR HMDs or allow faster selection. However, there is usually a tradeoff between smaller menu items and hence smaller layout on accuracy. Our research has not been focused on exploring the ideal size of menu items and this could be a possible line of research to help us develop techniques that require Hand or Hybrid selection of items.

There are several paths to further strengthen DMove. (1) The levels in one direction can be increased to allow for more items. This may be useful because, although the number of items in the radial menu is large enough to meet the needs of applications in AR systems, there can be cases which a large number of items are needed. As such, having more levels will allow more items to be included. (2) It is possible to optimize the layout further—e.g. finding the most suitable distance for each level in one direction instead of pre-defined values (i.e. 30cm) that we used in our study. (3) Since we want DMove to be accessed on-demand, future work can also focus on exploring ways to separate DMove from ordinary moving. We have done some preliminary explorations and one way that is possible for all commercial AR HMDs, for instance, is to use the third dimension (Y-axis) where users can perform an on tiptoe (up/down) action to wake up the DMove. This way, DMove can also be suitable for users with arm/hand disabilities as it does not require hands or any input device.

9 CONCLUSION

In this paper, we have presented DMove, a device-free and hands-free directional motion-based interaction for Augmented Reality (AR) Head-Mounted Displays (HMDs) that can be used for a range of applications including menu selection, remote control, and exergame. We first propose a method that can be used for recognizing directional movements in HMDs that does not need any additional external trackers. Then, we conduct a study to examine the accuracy of the proposed method for 8- and 16-block interfaces and also to understand their social acceptability and physical/mental comfort. We then optimize the interface based on findings from the first study and conduct a second study to compare the menu selection performance of DMove with Hand and Hybrid (Head+Hand) approaches.

We have found that (1) Our proposed recognition method is very accurate—100% for 8-block DMove and 98.06% accuracy for 16-block DMove; (2) Users prefer to use DMove in front of familiar people and indoor scenarios (like their home or office); (3) Users felt more discomfort when moving towards directions that they cannot see; (4) DMove has an equal task completion time as Hand and Hybrid and a lower error than Hand when using a current consumer AR HMD; and (5) DMove is preferred by users because it has low workload but high usability and novelty.

Based on our results, we list several design guidelines including allowing for customization due differences in users' physical features, placing frequently used items near the user and in the frontal directions, and offering visual and/or auditive feedback—no additional tactile feedback is needed because DMove inherently comes with it, as users can feel when their foot touches the ground.

ACKNOWLEDGMENTS

The authors would like to thank the anonymous reviewers for their valuable comments and helpful suggestions. The work is supported in part by Xi'an Jiaotong-Liverpool University (XJTLU) Key Special Fund (#KSF-A-03) and XJTLU Research Development Fund.

REFERENCES

[1] David Ahlström, Khalad Hasan, and Pourang Irani. 2014. Are you comfortable doing that?: acceptance studies of around-device gestures in and for public settings. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices & services - MobileHCI '14*. ACM Press, Toronto, ON, Canada, 193–202. https://doi.org/10.1145/2628363.2628381

[2] Jason Alexander, Teng Han, William Judd, Pourang Irani, and Sriram Subramanian. 2012. Putting your best foot forward: investigating real-world mappings for foot-based gestures. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, Austin, Texas, USA, 1229. https://doi.org/10.1145/2207676.2208575

[3] R. Atienza, R. Blonna, M. I. Saludares, J. Casimiro, and V. Fuentes. 2016. Interaction techniques using head gaze for virtual reality. In *2016 IEEE Region 10 Symposium (TENSYMP)*. 110–114. https://doi.org/10.1109/TENCONSpring.2016.7519387

[4] Ana M. Bernardos, David Gómez, and José R. Casar. 2016. A Comparison of Head Pose and Deictic Pointing Interaction Methods for Smart Environments. *International Journal of Human-Computer Interaction* 32, 4 (April 2016), 325–351. https://doi.org/10.1080/10447318.2016.1142054

[5] Aimee L. Betker, Ankur Desai, Cristabel Nett, Naaz Kapadia, and Tony Szturm. 2007. Game-based Exercises for Dynamic Short-Sitting Balance Rehabilitation of People With Chronic Spinal Cord and Traumatic Brain Injuries. *Physical Therapy* 87, 10 (Oct. 2007), 1389–1398. https://doi.org/10.2522/ptj.20060229

[6] Gunnar Borg. 1998. *Borg's perceived exertion and pain scales.* Human kinetics.

[7] Doug A. Bowman, Ernst Kruijff, Joseph J. LaViola, and Ivan Poupyrev. 2004. *3D User Interfaces: Theory and Practice.* Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA.

[8] Doug A. Bowman, Ryan P. McMahan, and Eric D. Ragan. 2012. Questioning naturalism in 3D user interfaces. *Commun. ACM* 55, 9 (Sept. 2012), 78. https://doi.org/10.1145/2330667.2330687

[9] Mihai Bâce, Teemu Leppänen, David Gil de Gomez, and Argenis Ramirez Gomez. 2016. ubiGaze: ubiquitous augmented reality messaging using gaze gestures. In *SIGGRAPH ASIA 2016 Mobile Graphics and Interactive Applications on - SA '16*. ACM Press, Macau, 1–5. https://doi.org/10.1145/2999508.2999530

[10] Rory M.S. Clifford, Nikita Mae B. Tuanquin, and Robert W. Lindeman. 2017. Jedi ForceExtension: Telekinesis as a Virtual Reality interaction metaphor. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, Los Angeles, CA, USA, 239–240. https://doi.org/10.1109/3DUI.2017.7893360

[11] Andrew Crossan, Stephen Brewster, and Alexander Ng. [n. d.]. Foot Tapping for Mobile Interaction. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*. 418–422.

[12] Heiko Drewes and Albrecht Schmidt. 2007. Interacting with the Computer Using Gaze Gestures. In *Human-Computer Interaction – INTERACT 2007*, Cécilia Baranauskas, Philippe Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa (Eds.). Vol. 4663. Springer Berlin Heidelberg, Berlin, Heidelberg, 475–488. https://doi.org/10.1007/978-3-540-74800-7_43

[13] Barrett Ens, Ahmad Byagowi, Teng Han, Juan David Hincapié-Ramos, and Pourang Irani. 2016. Combining Ring Input with Hand Tracking for Precise, Natural Interaction with Spatial Analytic Interfaces. In *Proceedings of the 2016 Symposium on Spatial User Interaction - SUI '16*. ACM Press, Tokyo, Japan, 99–102. https://doi.org/10.1145/2983310.2985757

[14] Kathrin Gerling, Ian Livingston, Lennart Nacke, and Regan Mandryk. 2012. Full-body motion-based game interaction for older adults. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. ACM Press, Austin, Texas, USA, 1873. https://doi.org/10.1145/2207676.2208324

[15] Peter J. Gianaros, Eric R. Muth, J. Toby Mordkoff, Max E. Levine, and Robert M. Stern. 2001. A Questionnaire for the Assessment of the Multiple Dimensions of Motion Sickness. *Aviation, space, and environmental medicine* 72, 2 (Feb. 2001), 115–119. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910410/

[16] David F. Goldspink. 2005. Ageing and activity: their effects on the functional reserve capacities of the heart and vascular smooth and skeletal muscles. *Ergonomics* 48, 11-14 (Nov. 2005), 1334–1351. https://doi.org/10.1080/00140130500101247

[17] Jun Gong, Yang Zhang, Xia Zhou, and Xing-Dong Yang. 2017. Pyro: Thumb-Tip Gesture Recognition Using Pyroelectric Infrared Sensing. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology - UIST '17*. ACM Press, Québec City, QC, Canada, 553–563. https://doi.org/10.1145/3126594.3126615

[18] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11, 1 (Nov. 2009), 10. https://doi.org/10.1145/1656274.1656278

[19] Teng Han, Jason Alexander, Abhijit Karnik, Pourang Irani, and Sriram Subramanian. 2011. Kick: investigating the use of kick gestures for mobile interactions. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services - MobileHCI '11*. ACM Press, Stockholm, Sweden, 29. https://doi.org/10.1145/2037373.2037379

[20] Sandra G. Hart. 1986. NASA Task load Index (TLX). Volume 1.0; Paper and pencil package. (1986).

[21] Khalad Hasan, David Ahlström, and Pourang Irani. 2013. Ad-binning: leveraging around device space for storing, browsing and retrieving mobile device content. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 899–908.

[22] Hamilton A. Hernandez, Zi Ye, T.C. Nicholas Graham, Darcy Fehlings, and Lauren Switzer. 2013. Designing action-based exergames for children with cerebral palsy. ACM Press, 1261. https://doi.org/10.1145/2470654.2466164

[23] Karen Ho and Hanley Weng. 2013. Favoured attributes of in-air gestures in the home environment. In *Proceedings of the 25th Australian Computer-Human Interaction Conference on Augmentation, Application, Innovation, Collaboration - OzCHI '13*. ACM Press, Adelaide, Australia, 171–174. https://doi.org/10.1145/2541016.2541095

[24] Daniel Horodniczy and Jeremy R. Cooperstock. 2017. Free the Hands! Enhanced Target Selection via a Variable-Friction Shoe. ACM Press, 255–259. https://doi.org/10.1145/3025453.3025625

[25] Aulikki Hyrskykari, Howell Istance, and Stephen Vickers. 2012. Gaze gestures or dwell-based interaction?. In *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12*. ACM Press, Santa Barbara, California, 229. https://doi.org/10.1145/2168556.2168602

[26] Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. 2010. Designing gaze gestures for gaming: an investigation of performance. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*. ACM Press, Austin, Texas, 323. https://doi.org/10.1145/1743666.1743740

[27] Robert J. K. Jacob. 1990. What you look at is what you get: eye movement-based interaction techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90*. ACM Press, Seattle, Washington, United States, 11–18. https://doi.org/10.1145/97243.97246

[28] Matthew J. Klein and Christina S. Simmers. 2009. Exergaming: virtual inspiration, real perspiration. *Young Consumers* 10, 1 (March 2009), 35–45. https://doi.org/10.1108/17473610910940774

[29] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–14. https://doi.org/10.1145/3173574.3173655

[30] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Vol. 5298. Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76. https://doi.org/10.1007/978-3-540-89350-9_6

[31] Magic Leap. 2018. Magic Leap One: Creator Edition | Magic Leap. https://www.magicleap.com/magic-leap-one

[32] Frank Chun Yat Li, David Dearman, and Khai N Truong. 2009. Virtual shelves: interactions with orientation aware devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*. ACM, 125–128.

[33] Chiuhsiang Joe Lin, Sui-Hua Ho, and Yan-Jyun Chen. 2015. An investigation of pointing postures in a 3D stereoscopic environment. *Applied Ergonomics* 48 (May 2015), 154–163. https://doi.org/10.1016/j.apergo.2014.12.001

[34] S. Malik, C. McDonald, and G. Roth. 2002. Hand tracking for interactive pattern-based augmented reality. In *Proceedings. International Symposium on Mixed and Augmented Reality*. IEEE Comput. Soc, Darmstadt, Germany, 117–126. https://doi.org/10.1109/ISMAR.2002.1115080

[35] Meta. 2018. Design - Meta User Interface Guidelines | Optimize Your Workspace. https://devcenter.metavision.com/design/user-interface-guidelines-optimize-your-workspace

[36] Meta. 2018. Meta | Augmented Reality. https://www.metavision.com

[37] Microsoft. 2018. Microsoft HoloLens. https://www.microsoft.com/en-us/hololens

[38] Mark R. Mine. 1995. *Virtual Environment Interaction Techniques*. TR95-018:507242-507248. University of North Carolina at Chapel Hill, Chapel Hill, NC, USA.

[39] Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*. ACM Press, Vancouver, BC, Canada, 177. https://doi.org/10.1145/1978942.1978969

[40] Joseph A Paradiso, Kai-Yuh Hsiao, and Ari Benbasat. [n. d.]. Interfacing to the Foot: Apparatus and Applications. In *CHI '00 Extended Abstracts on Human Factors in Computing Systems*. 2. https://doi.org/10.1145/633292.633389

[41] Hyung Min Park, Seok Han Lee, and Jong Soo Choi. 2008. Wearable augmented reality system using gaze interaction. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, Cambridge, UK, 175–176. https://doi.org/10.1109/ISMAR.2008.4637353

[42] Kun Qian, Chenshu Wu, Zimu Zhou, Yue Zheng, Zheng Yang, and Yunhao Liu. 2017. Inferring Motion Direction using Commodity Wi-Fi for Interactive Exergames. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems - CHI '17*. ACM Press, Denver, Colorado, USA, 1961–1972. https://doi.org/10.1145/3025453.3025678

[43] Mahfuz Rahman, Sean Gustafson, Pourang Irani, and Sriram Subramanian. 2009. Tilt techniques: investigating the dexterity of wrist-based input. In *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*. ACM Press, Boston, MA, USA, 1943. https://doi.org/10.1145/1518701.1518997

[44] Christian Reimann and Dirk Stichling. 2004. Kick-Up Menus. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems*. 1552–1552. https://doi.org/10.1145/985921.986128

[45] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Construction of a Benchmark for the User Experience Questionnaire (UEQ). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 4 (2017), 40. https://doi.org/10.9781/ijimai.2017.445

[46] Johannes Schöning, Florian Daiber, Antonio Krüger, and Michael Rohs. 2009. Using hands and feet to navigate and manipulate spatial data. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems - CHI EA '09*. ACM Press, Boston, MA, USA, 4663. https://doi.org/10.1145/1520340.1520717

[47] Jeremy Scott, David Dearman, Koji Yatani, and Khai N. Truong. 2010. Sensing foot gestures from the pocket. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology - UIST '10*. ACM Press, New York, New York, USA, 199. https://doi.org/10.1145/1866029.1866063

[48] Mike Sheinin and Carl Gutwin. 2014. Exertion in the small: improving differentiation and expressiveness in sports games with physical controls. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*. ACM Press, Toronto, Ontario, Canada, 1845–1854. https://doi.org/10.1145/2556288.2557385

[49] Linda E. Sibert and Robert J. K. Jacob. 2000. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '00*. ACM Press, The Hague, The Netherlands, 281–288. https://doi.org/10.1145/332040.332445

[50] Sanna E. Sihvonen, Sarianna Sipilä, and Pertti A. Era. 2004. Changes in Postural Balance in Frail Elderly Women during a 4-Week Visual Feedback Training: A Randomized Controlled Trial. *Gerontology* 50, 2 (2004), 87–95. https://doi.org/10.1159/000075559

[51] Marco Speicher, Anna Maria Feit, Pascal Ziegler, and Antonio Krüger. 2018. Selection-based Text Entry in Virtual Reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–13. https://doi.org/10.1145/3173574.3174221

[52] A. H. Taylor, N. T. Cable, G. Faulkner, M. Hillsdon, M. Narici, and A. K. Van Der Bij. 2004. Physical activity and older adults: a review of health benefits and the effectiveness of interventions. *Journal of Sports Sciences* 22, 8 (Aug. 2004), 703–725. https://doi.org/10.1080/02640410410001712421

[53] Franco Tecchia, Giovanni Avveduto, Raffaello Brondi, Marcello Carrozzino, Massimo Bergamasco, and Leila Alem. 2014. I'm in VR!: using your own hands in a fully immersive MR system. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology - VRST '14*. ACM Press, Edinburgh, Scotland, 73–76. https://doi.org/10.1145/2671015.2671123

[54] Boris Velichkovsky, Andreas Sprenger, and Pieter Unema. 1997. Towards gaze-mediated interaction: Collecting solutions of the "Midas touch problem". In *Human-Computer Interaction INTERACT '97*, Steve Howard, Judy Hammond, and Gitte Lindgaard (Eds.). Springer US, Boston, MA, 509–516. https://doi.org/10.1007/978-0-387-35175-9_77

[55] Frank Weichert, Daniel Bachmann, Bartholomäus Rudak, and Denis Fisseler. 2013. Analysis of the Accuracy and Robustness of the Leap Motion Controller. *Sensors* 13, 5 (May 2013), 6380–6393. https://doi.org/10.3390/s130506380

[56] Ian Witten and Eibe Frank. 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier. https://doi.org/10.1016/C2009-0-19715-5

[57] Steven L Wolf, Huiman X Barnhart, Gary L Ellison, Carol E Coogler, and Atlanta FICSIT Group. 1997. The Effect of Tai Chi Quan and Computerized Balance Training on Postural Stability in Older Subjects. *Physical Therapy* 77, 4 (April 1997), 371–381. https://doi.org/10.1093/ptj/77.4.371

[58] Seline Wüest, Nunzio Alberto Borghese, Michele Pirovano, Renato Mainetti, Rolf van de Langenberg, and Eling D. de Bruin. 2014. Usability and Effects of an Exergame-Based Balance Training Program. *Games for Health Journal* 3, 2 (April 2014), 106–114. https://doi.org/10.1089/g4h.2013.0093

[59] Tetsuya Yamamoto, Masahiko Tsukamoto, and Tomoki Yoshihisa. 2008. Foot-Step Input Method for Operating Information Devices While Jogging. In *2008 International Symposium on Applications and the Internet*. IEEE, Turku, Finland, 173–176. https://doi.org/10.1109/SAINT.2008.97

[60] D. Yu, K. Fan, H. Zhang, D. Monteiro, W. Xu, and H. Liang. 2018. PizzaText: Text Entry for Virtual Reality Systems Using Dual Thumbsticks. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (Nov. 2018), 2927–2935. https://doi.org/10.1109/TVCG.2018.2868581