HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices

YUKANG YAN, CHUN YU*†, XIN YI, and YUANCHUN SHI, Tsinghua University, China

We propose HeadGesture, a hands-free input approach to interact with Head Mounted Display (HMD) devices. Using HeadGesture, users do not need to raise their arms to perform gestures or operate remote controllers in the air . Instead, they perform simple gestures with head movement to interact with the devices. In this way, users' hands are free to perform other tasks, e.g., taking notes or manipulating tools. This approach also reduces the hand occlusion of the field of view [11] and alleviates arm fatigue [7]. However, one main challenge for HeadGesture is to distinguish the defined gestures from unintentional movements. To generate intuitive gestures and address the issue of gesture recognition, we proceed through a process of *Exploration - Design - Implementation - Evaluation*. We first design the gesture set through experiments on gesture space exploration and gesture elicitation with users. Then, we implement algorithms to recognize the gestures, including gesture segmentation, data reformation and unification, feature extraction, and machine learning based classification. Finally, we evaluate user performance of HeadGesture in the target selection experiment and application tests. The results demonstrate that the performance of HeadGesture is comparable to mid-air hand gestures, measured by completion time. Additionally, users feel significantly less fatigue than when using hand gestures and can learn and remember the gestures easily. Based on these findings, we expect HeadGesture to be an efficient supplementary input approach for HMD devices.

CCS Concepts: • **Human-centered computing → Gestural input**;

Additional Key Words and Phrases: Gesture, Head Movement Interaction, Virtual Reality

ACM Reference Format:
Yukang Yan, Chun Yu, Xin Yi, and Yuanchun Shi. 2018. HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 4, Article 198 (December 2018), 23 pages. https://doi.org/10.1145/3287076

1 INTRODUCTION

As the popularity of Head Mounted Display (HMD) devices rapidly increases, improving the input in virtual reality (VR) and augmented reality (AR) is of great value. Current state-of-the-art head mounted AR (e.g. Hololens [4]) and VR (e.g. HTC Vive [2]) devices mostly require users to input via in-air hand gestures or operating controllers. However, there are a number of situations where users cannot use their hands as the actuator [20]. For example, user's hands are sometimes occupied, e.g., while writing notes with a pen or holding the handrail on an escalator. For users with limited arm mobility, they cannot interact with HMD devices easily because of the requirement for hand input. Additionally, mid-air hand input for an extended time may cause arm fatigue and

*This is the corresponding author
†Also with Beijing Key Lab of Networked Multimedia.

Authors' address: Yukang Yan, yyk15@mails.tsinghua.edu.cn; Chun Yu, chunyu@tsinghua.edu.cn; Xin Yi, yix15@mails.tsinghua.edu.cn; Yuanchun Shi, shiyc@tsinghua.edu.cn, Key Laboratory of Pervasive Computing, Ministry of Education, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China.

2474-9567/2018/12-ART198 $15.00
https://doi.org/10.1145/3287076

pain [12, 53] which severely influences the user experience. In these scenarios, to provide a supplementary input approach that supports hands-free interaction is of great importance.

To support hands-free interaction, most HMD devices have a voice control interface. Voice control is accurate and convenient when in a quiet environment. However, it may suffer in a noisy environment and cannot easily protect users' privacy in the public. Head movements can be also leveraged as an input channel to HMD devices, especially as head positions are tracked by the built-in sensors of most current HMD devices. Users are also familiar with performing gestures with head movements, e.g. to nod or shake head to communicate. So in this paper, we explore interactions with HMD devices using only intentional head movements. We propose HeadGesture, a head movement based input approach for HMD devices. Using HeadGesture, users perform simple gestures with the head to issue the control commands of the HMD devices. HeadGesture supports basic operations, including pointing, dragging, zooming in and out, scrolling up and down, and returning to the homepage. With these operations, users can interact with different applications on HMD devices.

However, the design of HeadGesture faces two major challenges. First, the gestures should be simple and intuitive, otherwise users have to make excessive effort to learn and remember them, and this will discourage users. Secondly, the gestures should be robustly recognizable and distinguishable from unintentional head movements. Users will move their heads without prescribed gestures in mind, e.g., looking around, so we need to avoid these movements from being classified as HeadGesture (false positives[54]).

To address these challenges, we followed the *Exploration-Design-Implementation-Evaluation* procedure. In the *Exploration* and *Design*, we first explored the space of usable head gestures with users. Then, we applied the participatory design paradigm [62] to elicit the head gesture set from the users. Users were invited to design a gesture set for nine basic operations on HMD devices. Finally, we summarized design goals and generated the final gesture set based on the results. With the final gesture set, we implemented an algorithm to classify HeadGestures and to filter out unintentional head movements. The algorithm consists of four parts: gesture segmentation, data reformation and unification, feature extraction, and classification with the SVM model. Finally, we evaluated the performance and user experience of HeadGesture through a comparison experiment and application tests. The results showed that the overall performance of HeadGesture was comparable to mid-air hand gestures; meanwhile HeadGesture caused less fatigue and were easy to learn and remember. We concluded that HeadGesture is an promising supplementary input approach for HMD devices when hand input is not available.

2 RELATED WORK

We first review related work on hands-free interaction techniques and head movement based interactions. Then we discuss the participatory design paradigm for the design of gesture set.

2.1 Hands-free Interaction

To achieve the goal of hands-free interaction, e.g. text entry [25] and navigation [21], speech based input techniques were developed. However, with these techniques, recognition accuracy is compromised when in a noisy environment[27], and can be inconvenient to interact via speech in a quiet public place.

In addition to speech input, body gestures were also studied to support hands-free interactions. WIP (walking in place) [50] is an immersive and effective method for virtual locomotion by step-like movements. Using this method, techniques [44, 55, 57] support hands-free navigation in the virtual environment by sensing the headset motion of the user. Another body of leaning techniques [8, 16, 17, 48, 56, 61] can also achieve hands-free navigation. The users leans their body to the direction that they want to navigate to, the posture is sensed by chairs [8, 48], balance boards [17, 61]. Or the users can tilt part of their body, e.g. the wrist [16] or head [56] in the direction to navigate. These techniques, however, only support navigation in virtual space. In this paper, HeadGesture is designed to support most basic operations for different applications.

2.2 Head Movement Based Interaction

From infancy, humans naturally use head gestures to convey messages [34]. The applications of head gestures as a mode of human-computer interaction [36] and wheelchair control [14, 33] have been studied intensively, especially for users with limited hand or arm mobility.

For able-bodied users, head movement is also a valuable interaction channel. Studies have shown that sensing head orientation and position can help the calibration of gaze interaction and promote the accuracy of gaze-based selection tasks [51, 52]. Head movements were also leveraged to control desktop cursors [22, 58] and mobile devices [15], by mapping the position of the head to the cursor. Head gestures were also proposed for performing discrete operations on the desktop [40, 45, 54] and HMD glasses [19, 64]. HeadTurn [45] enables users to adjust input numeric values by turning their heads left or right beyond the range threshold. HeadPager[54] enables users to turn pages in two directions by leaning their heads to the left or the right area. HeadNod [40] supports quick dialogue answering via a nod or shake of the head. Glassgesture [64] was the first work to leverage head gestures to achieve user authentication on AR headsets. Smoothmoves [19] requires users to follow the movement trajectory of the target with the head to select it, on AR headsets.

Besides the preceding techniques only leveraging head movements, previous research also studied to combine the use of gaze and head movements [31, 32, 35]. As gaze can reflect the focus and intention of the user [39], it has been naturally leveraged as an input method [30]. Additionally, gaze changes have a strong correlation with head movement [10], this can promote the recognition of head orientation and head gestures [38]. Studies [31, 32] also showed that by combining the use of gaze and head movement data, target selection techniques can achieve higher performance than using only one of them (faster than head pointing, more accurate than eye pointing).

2.3 Participatory Design

A key to the design of gesture-based interfaces is the mapping between gesture and the command, this quantifies the discoverability and learnability of the gesture [24, 60]. However, current gestures are often created to manage constraints such as robust recognition rather than intuitiveness and ease of use [37, 41].

To improve the learnability and memorability of gestures, the user defined method was first proposed by Wobbrock et al. [62] by designing gestures on an interactive surface. This method shows participants a command and a simulation of the effect of that command, and then asks participants to define a gesture to issue it. After all participants define their gestures, the gesture with the highest consensus is assigned to the presented command. In comparison with the pre-defined gestures, the gesture sets that are generated by this method are of higher preference by the users [42], with higher memorability [43] and are consistent with users' acquired experience. [62]. This method has been successfully applied in the gesture design of many different areas, including mobile interaction [49], smart TV interaction [59], virtual reality [63] and augmented reality [46]. In this paper, we applied the participatory design paradigm to generate the HeadGesture set. To the best of our knowledge, we are the first to design head movement based gestures through a participatory design process.

3 HEADGESTURE DESIGN

The HeadGesture Design consisted of two parts, the gesture space exploration and the gesture elicitation experiment. Different from typical elicitation studies, we added a gesture space exploration before the elicitation experiment. Because users were not as familiar with head gestures as with hand or touchscreen gestures, this process was to analyze the design space of head gestures and to help users think of more usable gestures. Another difference was that the design goal of HeadGesture was two-fold. We aimed to generate a set of head gestures that were not only intuitive and natural to users, but also to be recognized robustly by computers. So we encouraged the participants to consider both goals when designing the gestures.

3.1 Gesture Space Exploration

The goal of this process was to explore the design space of head gestures. We extracted the dimensions that should be considered when designing head gestures. The results of this process helped participants in the following experiment to better design gestures to meet the design goals. Before exploration, we first summarized the related work on head gesture interaction. Table 1 lists the features that related work frequently used in their gesture design. Then, we went through the exploration process with users to gain more design inspiration and complete the design space. By combining the results of these two parts, we report the taxonomy of head gestures, which guides the gesture design, the design inspiration and strategies to avoid false positives.

Table 1. The features of gestures that related work leveraged in the design of head gesture interaction.

Features	Explanations for the Features
Movement Tracking	Map the different head orientations to the different positions of the cursors, by ray casting metaphors. (Nouse [22], HeadTilting [15])
Rotation	Rotate the head to the left or the right, with a range threshold to avoid the false positives. (HeadTurn [45], HeadPager [54] and [29])
Leaning	Leaning the Head to the shoulders, with the head facing forward. (HeadTilting [15] and [33])
Drawing Shapes	Drawing special shapes (e.g., circle, triangle) using the head. (GlassGesture [64])
Existing Gestures	Leveraging the existing head gestures, e.g., nodding and shaking. (HeadNod [40])

3.1.1 Participants. In the participatory exploration, we recruited 16 participants from a local campus. The average age was 24.44 (SD = 1.90). Four participants were female. Ten participants had experience with AR or VR HMD devices. All participants were familiar with gesture interaction.

3.1.2 Task. Users' task was to propose usable head gestures that met the design goals, and to report their design inspiration and strategies. In this process, we did not provide a command set to map the gestures to. The users were instructed to propose the gestures they thought usable for HMD interaction. They could think of a suitable function for the proposed gesture, but we did not limit the functions. The purpose was to collect a wider range of usable gestures, so that in the following elicitation experiment, participants would have more options to map to the commands. While designing the gestures, users wore a Hololens headset so they could take into account the weight of the device. We also showed a cursor and its trajectory of recent 500 ms in the center of the headset display. The cursor trajectory was to help users observe the amplitude and direction of the head rotations and movements.

3.1.3 Exploration Process. Participants first put the Hololens headset on and confirmed that they could see the cursor trajectory in their view. Then we introduced the concept of using head gestures to interact with HMD devices. We used the example of HeadPager [54] to explain the relationship of the gesture (rotating the head) and its function (turning pages). Then we clarified that in this process, we did not constrain the functions of the gestures and they should think of usable head gestures that met the design goals, as many as possible. We introduced the features summarized in Table 1 and suggested participants to use the features in the gesture design. We also encouraged participants to think of more usable features and novel gestures. Thirty minutes were given for the gesture design. After that, we reviewed the gestures and interviewed their composers on the inspiration and design strategies. In total, we collected 210 HeadGesture instances. Based on the results, we summarized the *Gesture Taxonomy*, *Design Inspiration* and *Strategies for Determining False Positives*.

3.1.4 Gesture Taxonomy. We summarized the gestures and categorized them along four dimensions, *Movement*, *Trajectory*, *Flow* and *Nature*. Within each dimension are multiple categories, shown in Table 2. *Movement* dimension includes five basic categories which breaks down the head gesture into head rotations, translations and dwellings. *Trajectory* describes the spatial features of head movement trajectories. We found participants typically use directional movements, or draw shapes and characters. *Flow* describes the temporal features of head movements, which were related to the strategies of participants to avoid the false positives. In *Nature* dimension, we summarized the design inspiration of the gestures.

Table 2. The head gesture taxonomy that we summarized from the results. The dimensions include movement category, movement trajectory, gesture flow and nature of the design.

Movement	Lower or Raise	Lower or raise the head along x axis
	Tilt	Rotate the head along y axis
	Rotate	Rotate the head along z axis
	Stretch	Stretch the neck to different directions without rotating the head
	Dwell	Stop the head movement for a short duration
Trajectory	directional	Move the head to different directions
	Shape	Use head to draw geometrical shapes, e.g., circle
	Character	Use head to write characters or numbers
Flow	Delimiter	To perform a head gesture at the start and the end as the delimiter to switch the mode
	Repetition	Repeat a head gesture for more than one times
	Reverse	Perform a head gesture and then reverse it
Nature	Transfer	Use the head movement to mimic the hand gestures
	Existence	Use the head gestures that already exist, e.g., nodding
	Infrequent Actions	Use the head movements that were rarely performed in daily life
	Large amplitude	Enlarge the amplitude of daily head movements

3.1.5 Design Inspiration. We summarized two categories of the design inspiration to achieve high intuitiveness and simplicity of the gestures.

1 *Act like using hands*: Participants proposed head gestures that mimic the use the hands. For example, P4 proposed "to raise the head fast towards the upper right corner" to mimic throwing objects away with the right hand, and "to lower head to the direction of the ground fast" to mimic putting away books into the drawer with hands. P7 proposed "to lower and shake the head simultaneously" to mimic swiping the hands to clean the window. With these metaphors, the composers of these gestures expected them to be easy to learn and memorize, and interesting to perform.
2 *Transferring Daily Experience*: Inspired by *Existing Gestures*, participants proposed gestures that already existed in daily life. For example, P3 proposed the action of leaning the head to the shoulder, which is performed when we feel tired. Another example is moving the head forward, which we perform to approach

a target to see it in detail. The composers believed that these gestures would reduce the learning effort and have a high discoverability.

3.1.6 Strategies for Determining False Positives. We summarized five categories of strategy that participants applied in their gesture design, to distinguish these gestures from unintentional head movements. These strategies can also be applied in combinations.

1 *Infrequent Actions*: Participants proposed gestures that are unlikely to be performed unintentionally, e.g., "rotate the head to the leftmost position". The advantage of these gestures was the low possibility of users accidentally performing them. However, some of these gestures were not natural feeling human movements, and these were excluded from the final set.
2 *Repeat It Twice*: Participants proposed to repeat a simple action two or more times to avoid false positives, e.g., "nod twice", "turn left twice". They thought that the repetition would confirm the intention of the input and maintain the simplicity and naturalness of the gesture.
3 *Draw Strokes*: Using the head to draw strokes was also proposed, these included geometric shapes (e.g., triangle, circle), characters and numbers. As the strokes have specific meanings, the gestures could be easily followed and remembered. Also, the meaning traces of the head movements helped with recognition.
4 *Delimiter Gestures*: The mode switch paradigm was also proposed, where users first perform a delimiter gesture to trigger the gesture input mode. Following this delimiter, users completed the rest of the gesture and perform the delimiter again as the end. In this way, little constraints were placed on type and range of gesture after the mode switch. This method still required the delimiter itself to be robustly recognized, and took more time to complete.
5 *Forth and Back*: Participants proposed a simple action followed by a reverse movement to avoid false positives. Similar to *Repeat It Twice*, participants believed that the reverse pattern is a guarantee of being intentional input. The examples included "Rotate to the left and back", "Move forward and back".

3.2 Gesture Elicitation Experiment

After the gesture exploration, we conducted a gesture elicitation experiment. Different from typical elicitation experiments [46, 49, 59], we provided the results of the exploration as reference for the participants. We printed the reference sheet that listed the *taxonomy*, including the dimensions and categories, the *features* summarized from related work and their gestures as examples, and the *inspiration* and *strategies* that were proposed by participants. Participants consulted these references during the design process, so they had more options to select from and more resources to draw inspiration form. Another difference was that we added an additional requirement for gestures, which was to be easily recognized as an intentional input. We decided to encourage users to consider this recognition issue because they knew the unintentional head movements they may perform when interacting with HMD devices, and therefore they could balance the trade-off of gesture intuitiveness and recognition from their perspective.

3.2.1 Command Set. The command set for HeadGesture covered basic control operations. We referred to the command set of a state-of-the-art AR headset (Hololens [4]), which included nine basic commands: *Drag*, *Hold*, *Home* (return to the main menu), *Scroll up/down*, *Select*, *Double Tap*, *Zoom in/out*. These commands could support most interaction tasks, including navigation and object manipulation. In most systems, *Double Tap* was implemented as a repetition of *Select*. However, we decided to separate the design for *Select* and *Double Tap* in this experiment. In this way, participants were less constrained on their design of *Select* and did not need to take into account whether the gesture of *Select* could be easily repeated for two times. As *Select* is the most fundamental operation, to separate it from Double Tap was to guarantee the highest quality design.

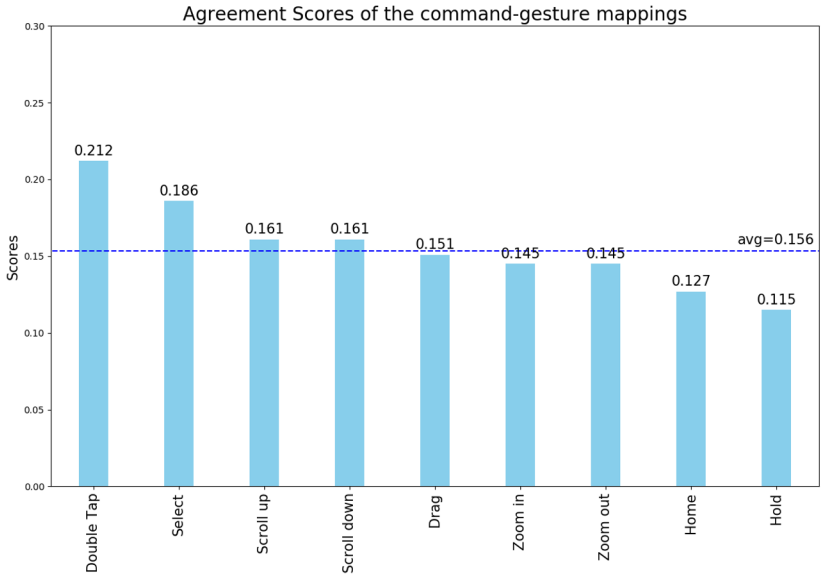

Fig. 1. The agreement scores for the head gestures that participants designed for each command.

3.2.2 Participants. We recruited sixteen participants for this study. Six were female. The average age was 25.56 (SD = 2.97). All of them were familiar with touchscreen gestures of smart phones. Four had experience with Hololens. Eight of them had also taken part in the exploration process.

3.2.3 Procedure. We followed the participatory design paradigm [62] to conduct this experiment. For each command, we first showed its effect on the Hololens by recorded screen videos. For example, we showed a video where the main menu of Hololens appeared in the center of the view to illustrate the effect of "Home" command. Participants wore the Hololens and watched the video, they were then asked to design a HeadGesture to trigger the command. During the design process, they were free to re-play the videos and free to pause and resume the videos. We encouraged participants to refer to the reference sheet, especially when they could not think of any suitable gestures. After they designed the gestures, they were instructed to perform the gestures along with videos, to examine whether the gesture matched the command. For each command, participants were asked to design more than one gesture. In this way, we had a wider range of gestures to choose from for the final set. To record the gestures, we implemented programs (at 60 frames per second) to record head orientation and position when participants performed gestures. Participants wrote down the explanations of the gestures and some of them also drew sketches to help explain them. The command order was randomized in this portion of the study.

3.2.4 Consensus Measurement. We collected 267 head gestures in total from 16 participants. The authors merged the same gestures manually based on the sketches and explanations of participants which resulted in 80 distinct gesture-command pairs. We applied a metric, *agreement score*, to measure the consensus of the command-gesture mappings of participants, first introduced by Wobbrock [62]. The results are shown by Figure 1. Compared to previous studies [46, 49, 59], where most scores were above 0.2, the agreement scores of the HeadGesture set were lower. This reflected that participants had less experience with head gestures and their proposals were less consistent. Another possible reason was that we allowed participants to propose more than one head gesture for each command. Comparing the agreement scores of different commands, we found that

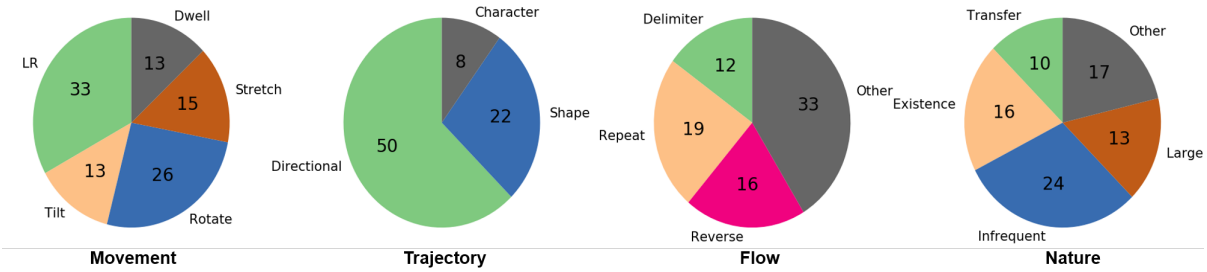

Fig. 2. The distribution of the gestures along the taxonomy dimensions that we generated in the exploration process. In *Movement*, the amount was over 80, because some of the features were used in combination. 'LR' represents 'Lower or Raise'.

participants were more consistent with the mappings for "Select" and "Double Tap". The proposals for "Hold" resulted in the highest disagreement. Most participants designed the gesture for "Double Tap" as a two-time repetition of the gesture for "Select", except two extra proposals for "Select". For all participants, the gestures proposed for "Scroll up/down" and "Zoom in/out" were in a reversed manner, so their agreement scores were equal. Table 3 lists the top three HeadGestures with the highest votes.

3.2.5 Gesture Distribution. Figure 2 shows the breakdown of the gestures collected in this experiment. We used the taxonomy that we generated in the exploration process to analyze the gestures. As shown, the most popular movements were *Lower or Raise* and *Rotate*, and they were also frequently used in combination, e.g., drawing circles with the head. *Directional* was the most frequent trajectory type, including directional rotations and stretching the neck to different directions. Directional rotations were reported to be simple and comfortable to perform. The amount of the gestures that applied *Delimiter*, *Repeat* and *Reverse* were similar. By applying these strategies, the designed gestures were more distinguishable from unintentional head movements. *Infrequent Actions* were the most frequent in nature, and gestures with *Transfer* nature were noted as interesting to perform.

3.2.6 Final Gesture Set. Based on participants' proposals, we generated the final gesture set. We first considered user preference and the selected gestures were all among the top three most voted proposals. However, we did not simply adopt the most popular proposals. We also ensured that there was no conflicts between the gestures and considered gesture practicality. The final gesture set is shown in Figure 3.

Select / Double Tap: We assigned *slow down, nod and back*, with highest votes, for *Select*, and a two-time repetition of *Select* to *Double Tap*. To select a target using HeadGesture, the user first controls the cursor to approach the target, stops the movement after entering the target boundary, and then perform the *nod and back* action to confirm the selection. The intentional stop, nod and back action help avoid false positives. Meanwhile, users are familiar with the *Nod* gesture from daily communication, so learning effort is minimal.

Drag: We modified the HeadGesture with the highest votes for *Drag*. We still used a delimiter gesture to switch the mode. But we applied the *Forth and Back* strategy and modified the delimiter gesture to be "Raise head and back to the front". Comparing to the original proposal, users could get their focus point back to the front after performing this HeadGesture, and could drag the objects in all directions. There was a metaphor for this HeadGesture, which was to load the object onto the head and to release it at the destination.

Hold: We adopted the HeadGesture "Quickly Rotate to the left and back". "Forth and Back" strategy and the fast speed helped avoid false positives. We suggested users to remember it as a nod to the left.

Home: We adopted participants' design of "Lean head to the shoulder", which received the highest number of votes. This gesture is hardly performed in daily life, except for when one feels tired or has neck pain. We regarded

Table 3. The list of top three head gestures that received the most votes for each command.

Operations	Most Frequent HeadGestures Proposed by Participants
Drag	Raise head in a large scale as a delimiter gesture.(7/26) Draw a circle using head as a delimiter gesture. (4/26) Move head forward/backward as a delimiter gesture. (3/26)
Hold	Move head forward and dwell. (5/33) Rotate head in one direction and back. (5/33) Draw a small circle in the front using head. (4/33)
Home	Lean head to one shoulder. (8/36) Circle head around. (5/36) Shake head once. (4/36)
Scroll up/down	Raise/Lower head to the highest/lowest position and dwell. (7/29, 7/29) Draw circles clockwise/anticlockwise using head, as if rotating a pulley. (5/29, 5/29) First lower/raise head and then raise/lower head in a large scale. (4/29, 4/29)
Select/Double Tap	Slow down, nod and back towards the forward direction.(7/26, 7/24) Draw a small circle in the front using head. (6/26, 6/24) Stretch the neck (head) forward and backward. (5/26, 5/24)
Zoom in/out	Stretch the neck forward/backward in a large scale and dwell. (9/32, 9/32) Raise/Lower head and back for twice. (4/32, 4/32) Lean head to the left/right shoulder. (4/32, 4/32)

this gesture to be suitable for the command "Home", because it was consistent with the concept of having a rest after interacting with an application for a period of time and one would want to return to the main menu.

Scroll up / down: For this command, we chose the HeadGesture with the second highest number of votes. We dropped the most popular proposal, because if someone performs *Scroll Up* several times in a row, the action was too similar to the delimiter of *Drag* and led to conflicts. Additionally, the focus point would be moved away, which may distract the user from her previous focus. Instead, the adopted proposal of drawing small circles could be distinguished from other HeadGestures and also kept the user's focus in the original region. The metaphor for this gesture was to draw small circles as if rotating a pulley to raise the page, and release it by rotating reversely.

Zoom in / out: We modified the HeadGesture with highest number of votes for *Zoom* command. Participants proposed to stretch the neck forward to zoom in, as if approaching to look at the content more clearly. We added a dwell after stretching the neck as confirmation to help the recognition.

Select, Double Tap, Drag, Hold require users to perform them inside the target. Users first move the cursor into the target, and then perform these HeadGestures to trigger the according functions, e.g., to select it for *Select*. As these HeadGestures all apply *Forth and Back* strategy, the recognition algorithm requires the cursor position of the start point and the end point to be inside the target, otherwise it will not trigger the function. For the other HeadGestures, anywhere users performing them can trigger the according functions.

4 IMPLEMENTATION

We implemented an algorithm for recognizing the HeadGestures and avoiding false positives. This recognition algorithm consisted of four phases. First, we segmented the potential HeadGestures from the continuous head movement stream. Then we reformed and unified the data into an appropriate format for feature extraction. After that, we extracted features from statistical characteristics and from the calculation of *Dynamic Time Warping* (DTW) algorithm. Finally, we used an SVM-based algorithm to classify the input data into one of the nine HeadGesture types or unintentional head movements.

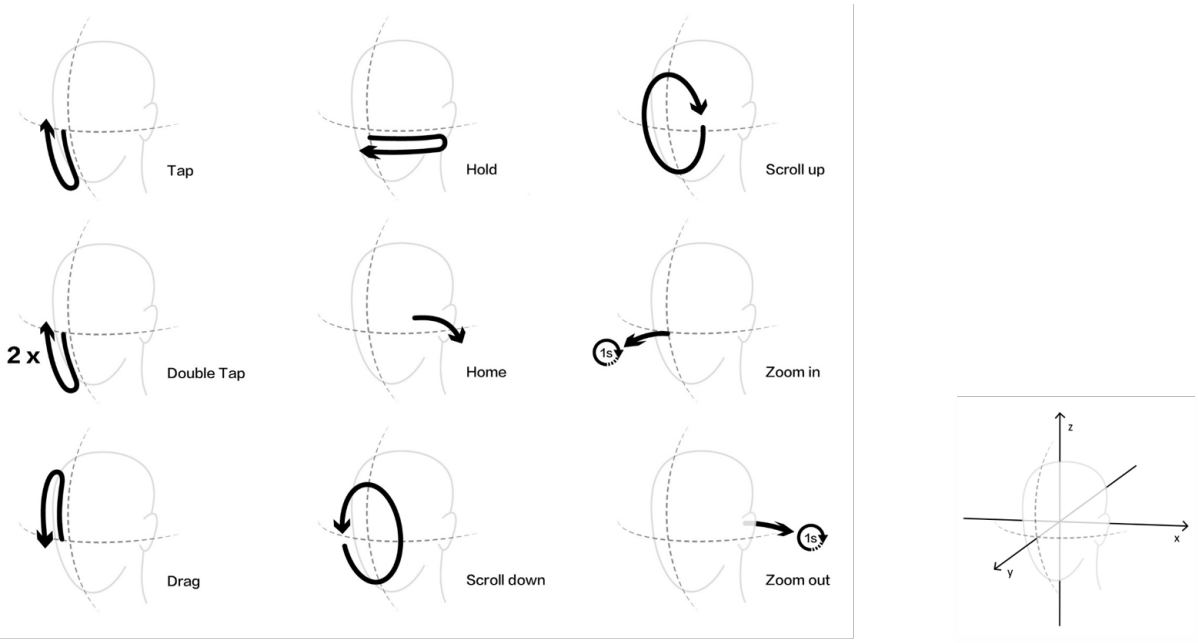

Fig. 3. The final design of the HeadGesture set for the nine commands. The movement of head is indicated by the arrows. "2×" represents the repeating of the action for twice. "1s" is an illustration for a dwell.

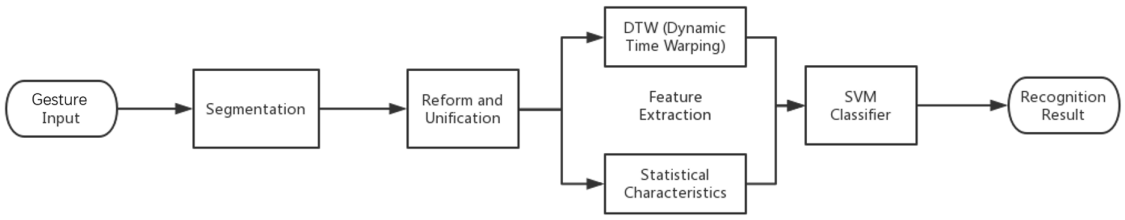

Fig. 4. The steps of the HeadGesture recognition. After the whole process, we classify the input head movement data into one of the HeadGesture types or the unintentional head movements.

4.1 Data Collection

Two types of data were collected for our machine learning algorithms, which were HeadGesture data and unintentional movement data. For the unintentional movement data, we designed representative tasks including walking around the room, browsing websites, and looking for targets which randomly appeared in 5m × 5m space. For the HeadGesture data, we asked users to perform the HeadGestures of the final set. We recorded their head movement data, including both head position and orientation.

We recruited ten participants from a local campus for this data collection process. Their average age was 22.54 (SD = 1.92) and four were female. Participants wore a Hololens headset to record head position and orientation. For the HeadGesture data, they performed each gesture for twenty times and the gestures were presented in a randomized order. For the unintentional movement data, we asked users to walk freely around the room, sit on a chair and browse news websites, and to look around to find the randomly placed targets which appeared every five seconds. Each of these unintentional tasks was performed for five minutes. Finally, we collected $20 \times 9 \times 10 = 1800$ HeadGestures, and $300 \times 3 \times 10 = 18000$ seconds of unintentional head movements.

4.2 Gesture Segmentation

This phase was to segment potential HeadGestures from the continuous stream of head movements. Head movement data was in the format of head position (x, y, z) and orientation (roll, pitch, yaw) data at 60 frames per second. As we observed, when users performed HeadGestures, they often started with an acceleration and ended with a deceleration. So we applied a simple segmentation algorithm, which required the head movement to start with the head rotation speed of 20 degrees per second, end with 4 degrees per second, and the whole duration should not exceed 2 seconds. For *Zoom in/out*, we set thresholds for head movement speed. We determined these parameters through a pilot test, which ensured all the HeadGestures that six participants performed could be extracted correctly. This segmentation was not overly strict, and allowed some unintentional head movements to pass. We would excluded these data in the following recognition phase.

4.3 Reform and Unification

This phase received the segmented data, then the data were reformed and unified to the required format for feature extraction. The first frame of the data determined the starting head position and orientation. To assist users in triggering the gestures at different head positions and orientations, we reformed the data from absolute coordinates to coordinates relative to this first frame. We set the head position and orientation of the first frame to be the point of origin, and calculated other frame data to subtract its value. We also calculated movement and rotation speeds and accelerations based on the position data, we smoothed them using the mean filter and median filter together[5]. After that, we unified the position and orientation data. Unified data enabled the following DTW algorithm to calculate the similarities and distances of the movement paths independently from the absolute rotation speed[13]. We unified the value of each position dimension from (-0.25m, 0.25m) to (0, 1), and each dimension of orientation (Euler angles [18]) from (-60 degrees, 90 degrees) to (0,1).

4.4 Feature Extraction

This phase was to extract spatial and temporal features from the input gesture data, and deliver these gestures to the following classifier. For spatial features, we first leveraged the head movement trajectory of the input data and calculated its similarity to the standard gesture templates. We applied Dynamic Time Warping (DTW) [9] to compute the similarities. The DTW algorithm received six channels of data, including the head position (x,y,z) and orientation (roll, pitch, yaw). Two of the authors performed the standard gestures to be the templates. For input data, we calculated its similarity to each template by the calculated *distance* between two trajectories. If the *distance* to the certain template satisfied the recognition threshold, we could infer that the input gesture belonged to that class. If the *distances* were all larger than the required range, the input gesture would be recognized as unintentional head movements. However, only using DTW was not sufficiently accurate. As we tested, within the HeadGestures, it resulted in the accuracy of 93.22% on average (SD = 1.7%). However, if we took the unintentional movements into consideration, the average accuracy dropped to 90.63%.

To improve recognition, we decided to use the results of the DTW algorithm as the basic features for how templates resemble input gestures [13] and add more statistical features to better characterize the HeadGestures.

The statistical features were extracted based on our observation of the HeadGesture behaviors. As some of the HeadGestures applied *Forth and Back* strategy, the offset of the start position and end position of the user's head should be small. So we calculated this *displacement* as another spatial feature. As some of the HeadGestures were directional rotations or drawing shapes, which consisted of sharp curves in the trajectory, we calculated the maximum, average value of *curvature* of the trajectory. The curvature was calculated only in the 2D x-y space of the head rotations. *Movement_distance* was also measured. For temporal features, we calculated the *maximum_speed, average_speed, maximum_acceleration, continuous_acceleration*. We calculated the features of accelerations because some HeadGestures required continuous speed up, like *Select* and *Drag*. The detailed calculations are listed below. We combined these features to be the input vector for the following SVM classifier.

$$path\_similarity = vector\ of\ [DTW(path_n, input\_gesture)\ for\ path_n\ in\ templates] \tag{1}$$

$$movement\_distance = \sum_{n=start+1}^{end} \|Position_n - Position_{n-1})\| \tag{2}$$

$$displacement = \|\ (Position_{end} - Position_{start})\| \tag{3}$$

$$maximum\_curvature = \max_{\forall n \in [start+1, end]} \frac{\left\|x_n^{'} \times y_n^{''} - y_n^{'} \times x_n^{''}\right\|}{(x_n^{'2} + y_n^{'2})^{\frac{3}{2}}} \tag{4}$$

$$average\_curvature = \frac{1}{end - start} \sum_{n=start+1}^{end} \frac{\left\|x_n^{'} \times y_n^{''} - y_n^{'} \times x_n^{''}\right\|}{(x_n^{'2} + y_n^{'2})^{\frac{3}{2}}} \tag{5}$$

$$maximum\_speed = \max_{\forall n \in [start+1, end]} \|Position_n - Position_{n-1}\| \tag{6}$$

$$average\_speed = \frac{1}{end - start} \sum_{n=start+1}^{end} \|Position_n - Position_{n-1}\| \tag{7}$$

$$maximum\_acceleration = \max_{\forall n \in [start+2, end]} (\|Position_n - Position_{n-1}\| - \|Position_{n-1} - Position_{n-2}\|) \tag{8}$$

$$continuous\_acceleration = \max_{\forall i, j \in [start+2, end]} \|j - i\|\ \ s.t.\ \forall n \in [i, j]\ speed_n > speed_{n-1} \tag{9}$$

4.5 Classification by SVM

We implemented a linear SVM classifier, with a linear kernel and the decision function type of "OneVSOne". This algorithm classified the input gesture into ten categories (nine HeadGestures and unintentional movements). It took the feature vector generated in the feature extraction phase as the input and calculated the most possible category as the output. We performed five-fold cross validation to evaluate its performance. We balanced the data by setting the training weight of unintentional movement data to be 0.1, as the unintentional data was roughly ten times of HeadGesture data. The performance of classification was 97.42% (SD = 1.12%) on average. In comparison with the performance of only using the DTW algorithm, the accuracy was much improved, which showed that the statistical features did improve the recognition.

5 EVALUATION

We evaluated the performance of HeadGesture as a mode of interaction with the HMD device. As target selection is the most frequent and fundamental interaction in AR and VR [47], we conducted a comparison experiment to compare the selection performance of using HeadGesture ("Select" gesture) to a mid-air hand gesture ("Air Tap" gesture of Hololens). For the other HeadGestures, we evaluated their performance through different application tasks. We measured the completion time and the subjective feedback of users.

5.1 Comparison Experiment

We first compared the performance of HeadGesture and the hand gesture in the target selection task. The selection task required users to control a cursor by head movements, which was fixed in the center of the view. Two approaches (HeadGesture and hand gesture) were used as confirmation of the selection after participants moved the cursor into the target. We measured the completion time and selection accuracy of the tasks and recorded users' comments. The experiment was split into three sections to test the learning effects of both approaches.

5.1.1 Participants. We recruited twelve participants from a local campus. To test the learning effect, we ensured that they had not participated in the *HeadGesture Design* and had no experience of using an AR headset. The average age was 23.58 (SD = 1.44). Four participants were female. Five participants reported they experienced motion sickness [26] when they watched 3D movies.

5.1.2 Apparatus. We used Hololens as the experiment platform and developed our own target selection application using the Unity 5 engine. The Hololens had a field of view of about 30° × 17.5° and sensing accuracy of about 2cm position error in translation and 2° in rotation. We developed a program to receive the head position and orientation data from the headset, through the standard HTTP protocol. We applied the classifier as introduced in *Implementation* for HeadGesture recognition.

5.1.3 Task. Participants selected the targets with a cursor. The cursor was fixed in the center of the field of view and followed the head movement of participants. A successful selection required the user to move the cursor into the target and perform a gesture, either HeadGesture or a hand gesture, inside the boundary. If the action occurred outside the target or was not recognized, it was recorded as a miss and the target needed to be selected again. The required HeadGesture was *Select*, as shown in Figure 3, the hand gesture was the default gesture of Hololens, *Air Tap*. The size of the target was 0.1m × 0.1m × 0.1m cubes. The position of the target was randomized in a 2m × 2m × 5m space, 2m in the front of the user. If unintentional head movements were recognized as *Select*, participants reported this to the experimenter. Overall, participants performed 2 approaches × 3 sessions × 50 tasks = 300 selections. The three sessions were to test the learning effect of two approaches, with a five-minute break between the sessions. The order of the two approaches was counter-balanced.

5.1.4 Procedure. On arrival, the Hololens device was introduced to the participants. They went through an official tutorial to learn the gesture interactions with the Hololens. We emphasized the introduction of "Air Tap" gesture, as they needed to perform this in the experiment, which was explained as "Make a loose fist, point up the index finger, quickly tap it down and all the way back up". Then we played videos of users performing *Select* HeadGesture to help them understand this gesture. After that, the warm-up session was provided for them to practice selecting targets with two approaches. Participants were given ten minutes' practice time to perform both gestures to select the targets, and then see their results. They could also replay the tutorials or videos if needed. After participants confirmed that they were familiar with the two approaches, they started the formal experiment. After the participant's session was completed, they completed a questionnaire to reflect on their experience, in the aspects of fatigue, learning effort, sickness and the natural feeling of each approach, in five-point Likert scale. On average, each participant took about 50 minutes.

5.1.5 Results. In total, we collected 12 participants × 300 tasks = 3600 selections. To measure the performance, we calculated the completion time and selection accuracy. To study the learning effect of two approaches, we analyzed the difference in performance between sessions. We also summarized the comments and feedback participants gave about the interaction experience.

5.1.5.1 Speed and Accuracy. The overall results are shown in Figure 5. We conducted RM-ANOVA test on completion time of two approaches in three sessions. The results showed that there was a significant difference of

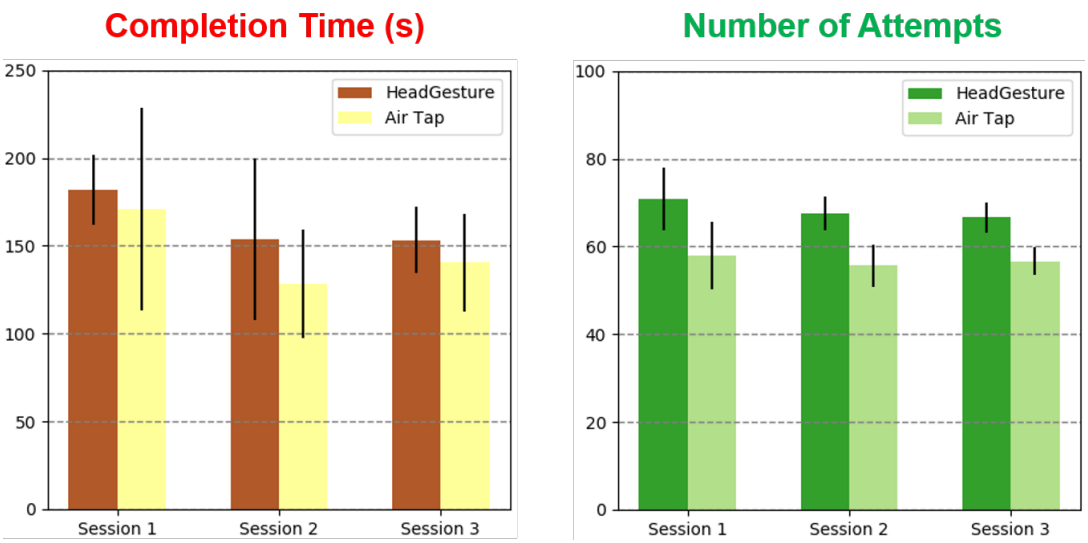

Fig. 5. The average completion time (left) and number of attempts (right) to complete 50 tasks in each session. The error bars represent standard deviation.

completion time between sessions ($F_{2,22} = 16.544$, $p < 0.002$), but no significant difference between two approaches ($F_{1,11} = 3.257$, $p = 0.099$). However, the average completion time of *Select* HeadGesture was higher than *Air Tap* in all three sessions. This was partially because the HeadGesture was less accurate, and required more attempts. Another reason was that selections were performed continuously, without resets. The new target appeared directly after the previous was selected. As a result, in the sessions of *Air Tap*, most participants kept their arms in the air across the selections until they felt tired. If resetting arm position after each selection was necessary, the completion time of *Air Tap* may have been longer. To confirm this effect, we calculated the selection time and the time of approaching the target separately. The average action time of *Select* HeadGesture was 0.45 seconds (STD = 0.12 seconds) after the segmentation. For *Air Tap*, we could not track the exact movement of the hand on Hololens, and therefore we regarded the time when the cursor entered the target as the starting moment of the selections. The average action time of *Air Tap* was 0.65 seconds (STD = 0.20 seconds). So *Air Tap* with arm position reset was even slower than *HeadGesture*. As the *Air Tap* selections in this experiment consisted of two types, which were selections with and without arm position resets, we added post tests to measure their speeds separately. The action time of *Air Tap* without resets was roughly 0.6 seconds and for *Air Tap* with resets, the time was about 1.5 seconds. So the selection time of *Select* HeadGesture was shorter, and longer completion time was caused by more attempts and the duration of adjustments after the failed attempts. We could expect, with improved recognition accuracy in the future, the completion time of HeadGesture could be further sped up.

For selection accuracy, we conducted RM-ANOVA test on the number of attempts of the two approaches in three sections. The result showed that *Select* HeadGesture made significantly more misses ($F_{1,11} = 57.539$, $p < 0.001$) than *Air Tap* and the difference between sessions were not significant ($F_{2,22} = 3.386$, $p = 0.052$). We found that overshoot frequently appeared in the misses, where the cursor was moved outside the target space. In the future, we could adjust the recognition mechanism to address this issue. For learning effect, the selection speed and accuracy of *Select* HeadGesture were improved session by session, meanwhile the best performance of *Air Tap* appeared at the second session. As participants reported, they felt tired in the final session even after the

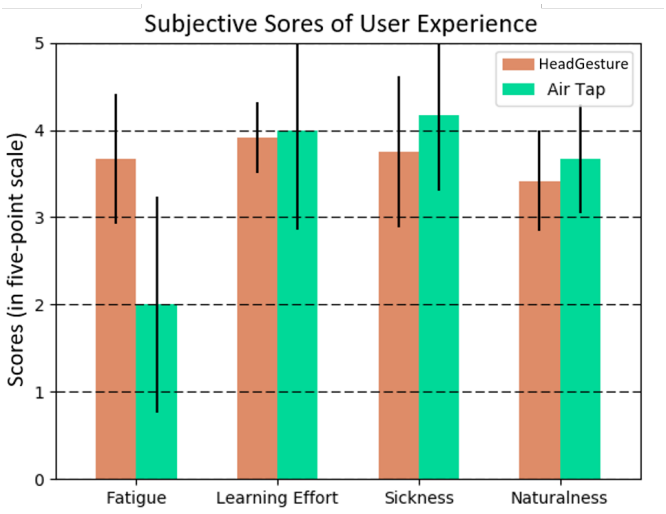

Fig. 6. The subjective scores that participants reported about their interaction experience, in five-point Likert scale. Higher scores represent more positive feelings. The error bars represent the standard deviations.

break, and they reset their arms more frequently between selections. The accumulated arm fatigue possibly led to the drop in performance.

5.1.5.2 Subjective Feedback. The subjective scores are shown in Figure 6. We included sickness in the subjective measurement, although motion sickness caused by visual-kinesthetic conflicts were observed less on AR headsets [6]. Our consideration was that participants need to rotate and move their heads frequently, especially in the sessions of HeadGesture, and we took into account that movements might aggravate visual conflicts and lead to sickness. We ran Wilcoxon test on the subjective scores, the results showed that using HeadGesture led to significant less fatigue ($Z = 3.58$, $p < 0.001$) than using *Air Tap*, but there was no significant difference for other dimensions ($Z = -0.09, -1.13, -0.45$, $p = 0.927, 0.257, 0.654$). As we expected, the average score of sickness for using HeadGesture was higher. However, as the scores showed, participants did not feel significant sickness using either approach (both averaged above 3.5). Meanwhile, using mid-air hand gestures caused significant arm fatigue (the average around 2), HeadGesture alleviated this problem.

Most of the participants thought two approaches to be simple, intuitive and easy to learn. P3 commented that *Select* HeadGesture was relaxing, and that he would use it if the input task was not urgent. Besides this, some participants thought that the experience of *Select* was smooth as they used their heads to both move the cursor and perform the selections, without switching to the hand. P4 reported that because head movements were also used to trigger selection, instead of only for navigation, he intentionally moved his head more carefully. This effect lowered his navigation speed to some degree. P6 was also concerned that if there were dense objects around the target, whether HeadGesture could still work effectively. We will test these issues in the future.

As recorded, there were 34 cases of false positives in total during the 3600 successful selections. The accuracy was higher than simulated, possibly because participants did not walk or move their bodies to a great extent during the experiment, and therefore head movements were more stable. So in the cases where users are sitting, e.g., working at a desk, the false positive rate would not be the main challenge. In the future, we will also test the performance of selection in unstable scenarios, e.g., when users walk or are on the bus.

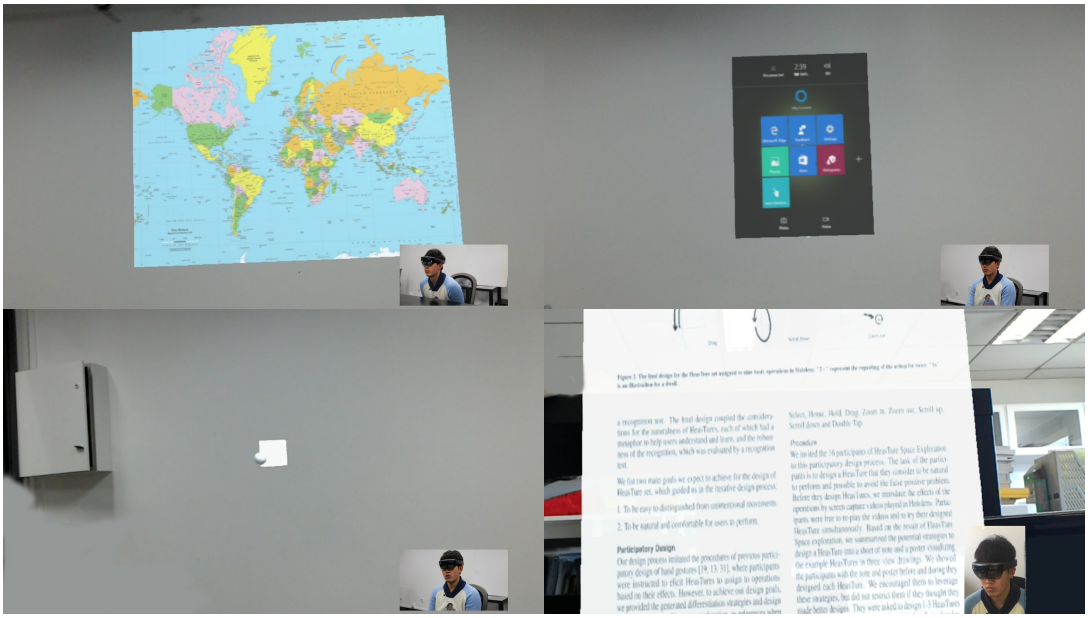

Fig. 7. Four applications that we developed to test the remaining HeadGestures. The tasks were map navigation (top left), menu invocation (top right), object translation (bottom left) and document reading (bottom right).

5.2 Application Test

For the remaining HeadGestures beyond *Select*, we developed four applications to test their performance in different tasks. The applications were typical interaction tasks on HMD devices, including map navigation, menu invocation, object translation and document reading. Through these tasks, we measured the completion time and collected the comments of participants.

5.2.1 Participants. We recruited six participants who had also participated in the comparison experiment, so that they were already familiar with interactions on HMD devices and the HeadGestures. Three of them were female. Their average age was 22.17 (SD = 1.17).

5.2.2 Tasks. As Figure 7 shows, we developed four applications: 1. Map navigation: the task was to find a specific country on the map, and participants were required to perform *Zoom in/out* to adjust the size of the map, and then performed *Hold* to confirm the selection; 2. Menu invocation: participants performed *Home* to invoke and close the main menu for ten times; 3. Object translation: participants controlled the cursor to approach the target object, and performed *Drag* to carry the object and then moved the object to the destination and performed another *Drag* to release it; 4. Document reading: participants read an article and performed *Scroll up/down* to turn page and back. Participants completed these applications in a random order. Before the tests, they were free to learn the HeadGestures and practice them. The experiment took about twenty minutes on average to complete. When the experiment was completed, we interviewed the participants to collect their comments and feedback.

5.2.3 Completion Time. Time to complete each HeadGesture was recorded. The results are presented in Table 4. The average time of performing each HeadGesture was less than one second. Currently on Hololens, to perform these tasks, users need to click the buttons on the top of the window to choose the operation (among "Zoom

in/out", "Drag", "Scroll up/down") and then approach the target and perform the selected operation on it. The entire process took much longer time compared to HeadGestures in our pilot test. This was because HeadGestures were expressive and saved time from switching to different modes. Within the HeadGestures, *Zoom in* took the longest time to perform. *Home* resulted in the highest standard deviations which suggested that participants made large difference in efficiency.

Table 4. The average, maximum completion time of different HeadGestures and the standard deviations.

	Double Tap	Drag	Hold	Home	Scroll Up	Scroll Down	Zoom In	Zoom Out
Avg (s)	0.62	0.49	0.51	0.77	0.78	0.79	0.94	0.84
Std (s)	0.10	0.15	0.18	0.30	0.29	0.25	0.26	0.28
Max (s)	0.75	0.68	0.74	1.25	1.19	1.30	1.35	1.28

5.2.4 Movement Range. Additionally, we recorded the movement ranges of performing different HeadGestures. The results of head rotation range are shown in Table 5. Because *Home* required participants to lean their heads, *Zoom In/Out* required them to move their heads without rotations, their movement ranges are not listed in the table. The average movement distances of *Zoom In/Out* gestures were 20.62 cm (std = 3.89cm) and 20.45 cm (std = 3.72cm). We calculated the average leaning angle of *Home* to be 55.05 °(std = 10.93 °). As shown, the rotation ranges of the HeadGestures were mostly less than 16 degrees, which were quite small and did not distract the participants' visual focus significantly. However, the movement of *Zoom In/Out* was about 20 cm long to ensure the robust recognition. We will try to shorten this distance in the future by improving the recognition algorithm.

Table 5. The rotation ranges of seven HeadGestures in angular coordinate systems, which are measured in alpha and phi dimensions separately. Alpha dimension is the rotation angle around z axis and phi is the rotation angle around x.

	Double Tap	Drag	Hold	Scroll Up	Scroll Down	Select
Alpha (°)	1.91	1.79	12.08	15.48	15.36	1.61
Std (°)	0.62	0.70	3.07	2.59	2.60	0.89
Phi (°)	10.12	10.43	2.19	12.38	12.46	9.19
Std (°)	1.29	1.64	1.36	2.21	2.34	1.78

5.2.5 Subjective Feedback. Most participants appreciated the design of the HeadGestures. For example, participants thought *Drag* and *Hold* to be simple and easy to perform. They reported that they could easily understand the metaphors of *Drag*, *Home*, *Scroll* and *Zoom*. They confirmed that these metaphors helped them memorize the HeadGestures. P1 commented the design of *Hold* and *Home* to be very interesting.

> "To load the object on my head was intuitive (*Hold*) and to rest the head onto the shoulder was interesting (*Home*)." - P1

In the object translation task, users needed to control the cursor to approach the target before performing *Drag*. Similar to *Select*, Some participants reported that to move the cursor into the boundary required high

concentration and felt it was tiring. A potential solution for this issue would be to apply the selection mechanism of the area cursor [23], which would reduce the requirement for selection.

In the document reading task, participants suggested implementing a continuous scroll mechanism. Currently, each time participants drew a complete circle using the head, the document would scroll with a fixed height (half a page). However, P1 and P3 suggested that the document could be scrolled in a continuous way as they rotated their heads. To support this, we could use *Scroll* as the delimiter to change the mode. The continuous scroll would be triggered after users complete the first circle and then followed the head rotation to scroll up or down.

> "It is cumbersome to perform *Zoom* several times in a row, considering that the movement distance is not short." - P5

For the design of *Zoom*, P5 argued the amplitude of the movement was too large to perform them in a row. We regarded this as a trade-off that the large movement helps us to distinguish it from unintentional movements but requires more effort of users. In the future, we will also test smaller amplitudes to optimize this trade-off. Besides this, most participants reported that when their hands were occupied, they would prefer to use HeadGestures.

6 DISCUSSION

Through the process of *Exploration-Design-Implementation-Evaluation*, we presented HeadGesture as a hands-free input approach for HMD control. Based on the results, we discuss the design implications for head movement based interactions, the applicable platform of HeadGestures, head orientation issue and the use case.

6.1 Design Implications

In *Design* and *Evaluation*, participants reported that they appreciated the designs illustrated by metaphors. The metaphors came from previous experience, e.g. nodding to confirm, or mirrored the behaviors of other channels, e.g. using the head to draw circles as if using the hand. With the metaphors, users could easily understand and remember the gestures and some users thought the interaction to be more interesting. The elicited metaphors could be applied to other tasks in the future. For example, *using head to swipe a window* can be mapped to command *Remove*, writing characters with heads can trigger the shortcuts of the applications. In *Exploration* and *Implementation*, we found that to distinguish HeadGestures from unintentional head movements, we needed to apply specific strategies including *infrequent movements, repetition of gestures, special strokes, delimiter gestures, forth and back*. By using these strategies, we proved that users could avoid most false positives in our evaluation. In *Design*, We used nine gestures in the final set, but there were more interesting gestures that were proposed by participants. Participants who like dancing proposed professional head gestures, e.g., the Head Slide [1], which was to move the head sideways without rotations. The advantages were that with very little possibility, these gestures would be performed unintentionally. However, it was not easy for regular users to perform. In the future, we can interview users with different backgrounds, e.g., professional dancers, to generate more usable head gestures for HMD interaction.

6.2 Applicable Platform

Although we implemented and evaluated HeadGesture on the platform of Hololens, we also tested it on VR platforms. With very little modification, our algorithm worked on VR headsets. We tested on HTC Vive headset, and all the HeadGestures were correctly recognized. The results showed that HeadGesture has good scalability and the potential to work on other HMD devices. The requirement of applying this approach is that the platform can track the head position and orientation, e.g. by lighthouse technique [28]. The tracking accuracy may influence the recognition performance and limit the amount of the HeadGestures that it can support. Besides HMD devices, HeadGesture also has the potential to be applied on other mobile platforms, like iPhone X [3] which can track the movement of the user's face. In that scenario, HeadGesture may support one-hand manipulation of smartphones.

6.3 Head Orientation

As explained in the *Implementation*, we reformed the absolute gesture data (head positions, orientations) into the relative data to the start frame. In this way, we enable to trigger the HeadGestures at different starting positions and orientations. In the *Evaluation*, users performed *Select*, *Drag*, *Zoom* and *Hold* at different head orientations, as the targets appeared at different positions. The classification results proved that these gestures can be correctly recognized after the data reformation. However, in the tasks of menu invocation and document reading, users mostly faced to the front and performed the HeadGestures in the same directions. For the HeadGestures evaluated in these tasks, we tested performing them in different directions after the *Evaluation*. Six users successfully triggered *Home* and *Scroll up/down* in nine different directions (facing to the left, right, up, down, front and four corners). *Home* could only be performed when users faced to the front, and they found it uncomfortable to lean their heads at other directions. The results inspired us, besides the recognition issue, the comfort level of HeadGestures may be different when performing them at different head orientations and it is of great value to further study this factor in the future.

6.4 Use Case

HeadGesture is to supplement hand operations for HMD devices, when users' hands are occupied or they feel fatigue, for example, when they wear a headset to read documents, and they want to write notes or draw sketches while reading. In situations analogous to this, to change the content or to select items in the view, their options will be to drop the pen and raise their hands to perform an in-air gesture, or to use HeadGesture. Using HeadGestures in these cases can help save time and reduce focus switches. As we interviewed participants in *Evaluation*, they consistently preferred HeadGesture when their hands were occupied or the current task was not urgent. However, as tested in *Evaluation*, the time efficiency of HeadGesture did not outperform hand gestures. So it is better not to use HeadGesture in the tasks that required continuous operations, e.g., text entry. For these tasks, users can still use hand input or controllers. Additionally, HeadGesture has the potential to be leveraged by users with limited arm mobility, and can improve their accessibility to the AR and VR headsets.

7 LIMITATION

In this paper, we leave some factors of HeadGesture to be studied in the future. One factor is the vocabulary size of HeadGesture. We designed nine different gestures to support the basic operations in HMD devices. However, the maximum capacity of gestures without conflicts was not tested in this paper. The second factor is the performance of HeadGesture in the daily living environment. Our evaluation experiment was conducted in the lab environment, and we need to test it in the more realistic environments, e.g. on the railway. In these scenarios, users' body will be more unstable and cause unpredictable head movements such as jitters. We will test whether the algorithm can still recognize the HeadGestures correctly under these circumstances. Another factor is different participant backgrounds. In this study, we recruited participants from the local campus. In the future, we could invite users with different cultures to generate more head gestures. For example, we could invite professional dancers to collect more interesting gestures, and consult the professional gesture designers for suggestions for avoiding false positives. The last one is the balance between movement range and recognition accuracy. Currently, some of the HeadGestures required relatively large movement range to perform to achieve robust recognition. This will limit the use scenarios, as users could not perform them in the crowded environment, e.g., on the bus. In the future, we will optimize this factor and improve the recognition performance.

8 CONCLUSION

We propose the HeadGesture approach, with which users perform simple and intuitive gestures using their heads to accomplish basic tasks on HMD devices. This approach supports hands-free interaction, so it releases the users' hands to work on other tasks simultaneously, and reduces arm fatigue. Through the process of *Exploration-Design-Implementation-Evaluation*, we completed the design and development of this approach. We generated the inspiration and strategies of the gesture design, which will be helpful for other research of head movement based interaction; we developed a recognition algorithm with the accuracy of 97.4% of recognizing ten categories of gestures (including distinguishing unintentional movements); and we evaluated the user performance and subjective feedback of using HeadGesture. Users thought this approach was easy and interesting to perform. We concluded that HeadGesture could be a supplementary input approach for HMD devices.

ACKNOWLEDGMENTS

This work is supported by the Natural Science Foundation of China under Grant No. 61521002, No. 61572276 and No. 61672314, Tsinghua University Research Funding No. 20151080408, and also supported by Beijing Key Lab of Networked Multimedia.

REFERENCES

[1] 2018. The Head Slide. Website. (2018). Retrieved July 22th, 2018 from https://www.youtube.com/watch?v=BH7fIPEPvoY.
[2] 2018. HTC Vive. Website. (2018). Retrieved March 7, 2018 from https://www.vive.com/us/.
[3] 2018. iPhone X. Website. (2018). Retrieved March 7, 2018 from https://www.apple.com/iphone-x/.
[4] 2018. Microsoft Hololens. Website. (2018). Retrieved March 7, 2018 from https://www.microsoft.com/en-us/hololens.
[5] GR Arce. 2005. Nonlinear Signal Processing: A Statistical Approach Wiley: New Jersey. (2005).
[6] Ronald T Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators & Virtual Environments* 6, 4 (1997), 355–385.
[7] Yuki Ban, Takuji Narumi, Tatsuya Fujii, Sho Sakurai, Jun Imura, Tomohiro Tanikawa, and Michitaka Hirose. 2013. Augmented Endurance: Controlling Fatigue While Handling Objects by Affecting Weight Perception Using Augmented Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 69–78. https://doi.org/10.1145/2470654.2470665
[8] Steffi Beckhaus, Kristopher J Blom, and Matthias Haringer. 2007. ChairIO–the chair-based Interface. *Concepts and technologies for pervasive games: a reader for pervasive gaming research* 1 (2007), 231–264.
[9] Donald J Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series.. In *KDD workshop*, Vol. 10. Seattle, WA, 359–370.
[10] B Biguer, M Jeannerod, and C Prablanc. 1982. The coordination of eye, head, and arm movements during reaching at a single visual target. *Experimental brain research* 46, 2 (1982), 301–304.
[11] Volkert Buchmann, Stephen Violich, Mark Billinghurst, and Andy Cockburn. 2004. FingARtips: Gesture Based Direct Manipulation in Augmented Reality. In *Proceedings of the 2Nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia (GRAPHITE '04)*. ACM, New York, NY, USA, 212–221. https://doi.org/10.1145/988834.988871
[12] Marcio C. Cabral, Carlos H. Morimoto, and Marcelo K. Zuffo. 2005. On the Usability of Gesture Interfaces in Virtual Reality Environments. In *Proceedings of the 2005 Latin American Conference on Human-computer Interaction (CLIHC '05)*. ACM, New York, NY, USA, 100–108. https://doi.org/10.1145/1111360.1111370
[13] Xiang 'Anthony' Chen and Yang Li. 2016. Bootstrapping User-Defined Body Tapping Recognition with Offline-Learned Probabilistic Representation. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 359–364. https://doi.org/10.1145/2984511.2984541
[14] Douglas A Craig and Hung T Nguyen. 2006. Wireless real-time head movement system using a personal digital assistant (PDA) for control of a power wheelchair. In *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 772–775.
[15] Andrew Crossan, Mark McGill, Stephen Brewster, and Roderick Murray-Smith. 2009. Head Tilting for Interaction in Mobile Contexts. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09)*. ACM, New York, NY, USA, Article 6, 10 pages. https://doi.org/10.1145/1613858.1613866
[16] Andrew Crossan, John Williamson, Stephen Brewster, and Rod Murray-Smith. 2008. Wrist Rotation for Interaction in Mobile Contexts. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '08)*. ACM, New York, NY, USA, 435–438. https://doi.org/10.1145/1409240.1409307

[17] Gerwin de Haan, Eric J. Griffith, and Frits H. Post. 2008. Using the Wii Balance Board&Trade; As a Low-cost VR Interaction Device. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology (VRST '08)*. ACM, New York, NY, USA, 289–290. https://doi.org/10.1145/1450579.1450657

[18] James Diebel. 2006. Representing attitude: Euler angles, unit quaternions, and rotation vectors. *Matrix* 58, 15-16 (2006), 1–35.

[19] Augusto Esteves, David Verweij, Liza Suraiya, Rasel Islam, Youryang Lee, and Ian Oakley. 2017. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology (UIST '17)*. ACM, New York, NY, USA, 167–178. https://doi.org/10.1145/3126594.3126616

[20] Marcela Fejtová, Luis Figueiredo, Petr Novák, Olga Štěpánková, and Ana Gomes. 2009. Hands-free interaction with a computer and other technologies. *Universal Access in the Information Society* 8, 4 (2009), 277.

[21] Jinjuan Feng and Andrew Sears. 2004. Using Confidence Scores to Improve Hands-free Speech Based Navigation in Continuous Dictation Systems. *ACM Trans. Comput.-Hum. Interact.* 11, 4 (Dec. 2004), 329–356. https://doi.org/10.1145/1035575.1035576

[22] Dmitry O Gorodnichy and Gerhard Roth. 2004. Nouse 'use your nose as a mouse' perceptual vision technology for hands-free games and interfaces. *Image and Vision Computing* 22, 12 (2004), 931–942.

[23] Tovi Grossman and Ravin Balakrishnan. 2005. The Bubble Cursor: Enhancing Target Acquisition by Dynamic Resizing of the Cursor's Activation Area. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 281–290. https://doi.org/10.1145/1054972.1055012

[24] Chris Harrison, Robert Xiao, Julia Schwarz, and Scott E. Hudson. 2014. TouchTools: Leveraging Familiarity and Skill with Physical Tools to Augment Touch Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2913–2916. https://doi.org/10.1145/2556288.2557012

[25] Jibo He, Alex Chaparro, Bobby Nguyen, Rondell Burge, Joseph Crandall, Barbara Chaparro, Rui Ni, and Shi Cao. 2013. Texting While Driving: Is Speech-based Texting Less Risky Than Handheld Texting?. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '13)*. ACM, New York, NY, USA, 124–130. https://doi.org/10.1145/2516540.2516560

[26] Lawrence J Hettinger and Gary E Riccio. 1992. Visually induced motion sickness in virtual environments. *Presence: Teleoperators & Virtual Environments* 1, 3 (1992), 306–310.

[27] Hans-Günter Hirsch and David Pearce. 2000. The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. In *ASR2000-Automatic Speech Recognition: Challenges for the new Millenium ISCA Tutorial and Research Workshop (ITRW)*.

[28] Gabriel M Hughes. [n. d.]. Moving Forward with Virtual Reality. *Spring and Summer 2015* ([n. d.]), 43.

[29] Anja Jackowski, Marion Gebhard, and Axel Gräser. 2016. A novel head gesture based interface for hands-free control of a robot. In *Medical Measurements and Applications (MeMeA), 2016 IEEE International Symposium on*. IEEE, 1–6.

[30] Rob Jacob and Sophie Stellmach. 2016. What You Look at is What You Get: Gaze-based User Interfaces. *interactions* 23, 5 (Aug. 2016), 62–65. https://doi.org/10.1145/2978577

[31] Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. 2015. MAGIC Pointing for Eyewear Computers. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers (ISWC '15)*. ACM, New York, NY, USA, 155–158. https://doi.org/10.1145/2802083.2802094

[32] Shahram Jalaliniya, Diako Mardanbeigi, Thomas Pederson, and Dan Witzner Hansen. 2014. Head and Eye Movement As Pointing Modalities for Eyewear Computers. In *Proceedings of the 2014 11th International Conference on Wearable and Implantable Body Sensor Networks Workshops (BSNWORKSHOPS '14)*. IEEE Computer Society, Washington, DC, USA, 50–53. https://doi.org/10.1109/BSN.Workshops.2014.14

[33] Pei Jia, Huosheng H Hu, Tao Lu, and Kui Yuan. 2007. Head gesture recognition for hands-free control of an intelligent wheelchair. *Industrial Robot: An International Journal* 34, 1 (2007), 60–68.

[34] Viktoria A Kettner and Jeremy IM Carpendale. 2013. Developing gestures for no and yes: Head shaking and nodding in infancy. *Gesture* 13, 2 (2013), 193–209.

[35] 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, New York, NY, USA, Article 81, 14 pages. https://doi.org/10.1145/3173574.3173655

[36] Edmund LoPresti, David M. Brienza, Jennifer Angelo, Lars Gilbertson, and Jonathan Sakai. 2000. Neck Range of Motion and Use of Computer Head Controls. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies (Assets '00)*. ACM, New York, NY, USA, 121–128. https://doi.org/10.1145/354324.354352

[37] Shahzad Malik, Abhishek Ranjan, and Ravin Balakrishnan. 2005. Interacting with Large Displays from a Distance with Vision-tracked Multi-finger Gestural Input. In *Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology (UIST '05)*. ACM, New York, NY, USA, 43–52. https://doi.org/10.1145/1095034.1095042

[38] Diako Mardanbegi, Dan Witzner Hansen, and Thomas Pederson. 2012. Eye-based Head Gestures. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 139–146. https://doi.org/10.1145/2168556.2168578

[39] Richard A Monty and John W Senders. 2017. *Eye movements and psychological processes*. Routledge.
[40] Louis-Philippe Morency and Trevor Darrell. 2006. Head Gesture Recognition in Intelligent Interfaces: The Role of Context in Improving Recognition. In *Proceedings of the 11th International Conference on Intelligent User Interfaces (IUI '06)*. ACM, New York, NY, USA, 32–38. https://doi.org/10.1145/1111449.1111464
[41] Meredith Ringel Morris, Anqi Huang, Andreas Paepcke, and Terry Winograd. 2006. Cooperative Gestures: Multi-user Gestural Interactions for Co-located Groupware. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1201–1210. https://doi.org/10.1145/1124772.1124952
[42] Meredith Ringel Morris, Jacob O Wobbrock, and Andrew D Wilson. 2010. Understanding users' preferences for surface gestures. In *Proceedings of graphics interface 2010*. Canadian Information Processing Society, 261–268.
[43] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-designed and User-defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1099–1108. https://doi.org/10.1145/2470654.2466142
[44] Niels Christian Nilsson, Stefania Serafin, and Rolf Nordahl. 2014. The Influence of Step Frequency on the Range of Perceptually Natural Visual Walking Speeds During Walking-in-place and Treadmill Locomotion. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology (VRST '14)*. ACM, New York, NY, USA, 187–190. https://doi.org/10.1145/2671015.2671113
[45] Tomi Nukarinen, Jari Kangas, Oleg Špakov, Poika Isokoski, Deepak Akkil, Jussi Rantala, and Roope Raisamo. 2016. Evaluation of HeadTurn: An Interaction Technique Using the Gaze and Head Turns. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 43, 8 pages. https://doi.org/10.1145/2971485.2971490
[46] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined gestures for augmented reality. In *IFIP Conference on Human-Computer Interaction*. Springer, 282–299.
[47] I. POUPYREV. 1998. Egocentric object manipulation in virtual environments : Empirical evaluation of interaction techniques. *Computer Graphics Forum, EUROGRAPHICS'98* 17, 3 (1998), 41–52. https://doi.org/10.1111/1467-8659.00252
[48] Kathrin Probst, David Lindlbauer, Michael Haller, Bernhard Schwartz, and Andreas Schrempf. 2014. A Chair As Ubiquitous Input Device: Exploring Semaphoric Chair Gestures for Focused and Peripheral Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 4097–4106. https://doi.org/10.1145/2556288.2557051
[49] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined Motion Gestures for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 197–206. https://doi.org/10.1145/1978942.1978971
[50] Mel Slater, Martin Usoh, and Anthony Steed. 1994. Steps and ladders in virtual reality. In *Virtual Reality Software And Technology*. World Scientific, 45–54.
[51] Oleg Špakov, Poika Isokoski, and Päivi Majaranta. 2014. Look and lean: accurate head-assisted eye pointing. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. ACM, 35–42.
[52] Oleg Špakov and Päivi Majaranta. 2012. Enhanced gaze interaction using simple head gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 705–710.
[53] Richard Stoakley, Matthew J. Conway, and Randy Pausch. 1995. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, 265–272. https://doi.org/10.1145/223904.223938
[54] Zhenyu Tang, Chenyu Yan, Sijie Ren, and Huagen Wan. 2016. HeadPager: Page Turning with Computer Vision Based Head Interaction. In *Asian Conference on Computer Vision*. Springer, 249–257.
[55] Sam Tregillus. 2016. VR-Drop: Exploring the Use of Walking-in-Place to Create Immersive VR Games. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 176–179. https://doi.org/10.1145/2851581.2890374
[56] Sam Tregillus, Majed Al Zayer, and Eelke Folmer. 2017. Handsfree Omnidirectional VR Navigation Using Head Tilt. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4063–4068. https://doi.org/10.1145/3025453.3025521
[57] Sam Tregillus and Eelke Folmer. 2016. VR-STEP: Walking-in-Place Using Inertial Sensing for Hands Free Navigation in Mobile VR Environments. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 1250–1255. https://doi.org/10.1145/2858036.2858084
[58] Javier Varona, Cristina Manresa-Yee, and Francisco J Perales. 2008. Hands-free vision-based interface for computer accessibility. *Journal of Network and Computer Applications* 31, 4 (2008), 357–374.
[59] Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control. In *Proceedings of the 10th European Conference on Interactive TV and Video (EuroITV '12)*. ACM, New York, NY, USA, 45–48. https://doi.org/10.1145/2325616.2325626
[60] Julie Wagner, Eric Lecolinet, and Ted Selker. 2014. Multi-finger Chords for Hand-held Tablets: Recognizable and Memorable. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2883–2892. https://doi.org/10.1145/2556288.2556958

[61] Jia Wang and Robert W. Lindeman. 2011. Silver Surfer: A System to Compare Isometric and Elastic Board Interfaces for Locomotion in VR. In *Proceedings of the 2011 IEEE Symposium on 3D User Interfaces (3DUI '11)*. IEEE Computer Society, Washington, DC, USA, 121–122. http://dl.acm.org/citation.cfm?id=2013881.2014229

[62] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[63] Yukang Yan, Chun Yu, Xiaojuan Ma, Xin Yi, Ke Sun, and Yuanchun Shi. 2018. VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 78, 13 pages. https://doi.org/10.1145/3173574.3173652

[64] Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li. 2016. Glassgesture: Exploring head gesture interface of smart glasses. In *Computer Communications, IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on*. IEEE, 1–9.

Received May 2018; revised August 2018; accepted October 2018