Classifying Head Movements to Separate Head-Gaze and Head Gestures as Distinct Modes of Input

Baosheng James Hou
Lancaster University
Lancaster, United Kingdom
b.hou2@lancaster.ac.uk

Joshua Newn
Lancaster University
Lancaster, United Kingdom
j.newn@lancaster.ac.uk

Ludwig Sidenmark
Lancaster University
Lancaster, United Kingdom
l.sidenmark@lancaster.ac.uk

Anam Ahmad Khan
National University of Science and Technology
Islamabad, Pakistan
12bscsaakhan@seecs.edu.pk

Per Bækgaard
Technical University of Denmark
Kgs. Lyngby, Denmark
pgba@dtu.dk

Hans Gellersen
Lancaster University
Lancaster, United Kingdom
Aarhus University
Aarhus, Denmark
h.gellersen@lancaster.ac.uk

ABSTRACT

Head movement is widely used as a uniform type of input for human-computer interaction. However, there are fundamental differences between head movements coupled with gaze in support of our visual system, and head movements performed as gestural expression. Both *Head-Gaze* and *Head Gestures* are of utility for interaction but differ in their affordances. To facilitate the treatment of Head-Gaze and Head Gestures as separate types of input, we developed HeadBoost as a novel classifier, achieving high accuracy in classifying gaze-driven versus gestural head movement (F_1-Score: 0.89). We demonstrate the utility of the classifier with three applications: gestural input while avoiding unintentional input by Head-Gaze; target selection with Head-Gaze while avoiding Midas Touch by head gestures; and switching of cursor control between Head-Gaze for fast positioning and Head Gesture for refinement. The classification of Head-Gaze and Head Gesture allows for seamless head-based interaction while avoiding false activation.

CCS CONCEPTS

• **Human-centered computing** → **Gestural input**; **Virtual reality**;

KEYWORDS

Head Gestures, Eye Tracking, Virtual Reality, Eye-head Coordination, Computational Interaction, Machine Learning, XGBoost

ACM Reference Format:

Baosheng James Hou, Joshua Newn, Ludwig Sidenmark, Anam Ahmad Khan, Per Bækgaard, and Hans Gellersen. 2023. Classifying Head Movements to Separate Head-Gaze and Head Gestures as Distinct Modes of Input. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3544548.3581201

CHI '23, April 23–28, 2023, Hamburg, Germany

ACM ISBN 978-1-4503-9421-5/23/04…$15.00
https://doi.org/10.1145/3544548.3581201

1 INTRODUCTION

Head movement is compelling as an input modality as it affords hands-free and non-verbal interaction. Head orientation is an implicit cue for attention, and an approximation of where we look [49]. We can move our heads quickly over a wide motion range, and have more control over head movement than over gaze for precise input [29, 44]. We are also expressive with head movement and use head gestures in everyday non-verbal communication [20]. Based on these properties, head movement has been leveraged for diverse purposes in interaction with computers, including viewpoint control and adaptive displays [34, 50, 52], pointing at desktops and in 3D [39, 45], and gestures to confirm input and issue commands [53, 54]. While head movement has proven so versatile, it has been treated as a uniform type of input, tracked as a stream of coordinates, and not further qualified.

In this work, we propose to distinguish between head movements that are gaze-driven, and head movements that are expressions of the head in its own right. We refer to the two types of movement as *Head-Gaze* versus *Head Gestures*. The dichotomy is fundamental as it is grounded in how head movement and the visual system interact. Much of our head movement is completely driven by gaze behaviour, to support the eyes in directing visual attention to objects that are not already right in front of us, and to keep the eyes within a comfortable eye-in-head position [30, 44]. In contrast, when we move our heads independently of gaze, to describe a gesture, then it is the head that leads the interaction and the visual system that adapts, making it possible to use head movements as means of expression without disrupting vision.

Both types of head movement are of utility for human-computer interaction (HCI), Head-Gaze as it approximates where we look, and Head Gestures as they afford more control and expressiveness. However, as current head-tracked interfaces treat all head movement as the same, head gestures have to be designed to avoid inadvertent trigger by Head-Gaze, and avoided altogether when head movement is used as proxy for gaze [12, 54, 55]. These problems can be overcome if Head-Gaze and Head Gestures are treated as distinct modes of input, while also enabling their use in tandem.

Our principal aim in this work is to classify head movements into Head-Gaze and Head Gestures, in order to enable their use as distinct modes of input. A previous work, *BimodalGaze*, explored refinement of gaze input with head gestures and proposed a heuristic for switching between a gaze mode and head mode, to avoid unintended input from head movement associated with the initial gaze shift [46]. The work demonstrated the complexity of separating Head-Gaze from Head Gestures. When the head supports gaze, it moves more slowly than the eyes, with most of the head movement typically occurring while gaze is already on target. As the head catches up, the eyes compensate for the head movement to fixate on the target. The same relative movement of head and eye is observed during head gestures, where the eyes compensate head movement to stabilise vision, making it difficult to distinguish gestural from gaze-driven behaviour. To overcome limitations observed with a heuristic approach, we explore machine learning for the classification of Head-Gaze versus Head Gestures.

The first problem we address is producing a data set to train a classifier, as existing data sets of eye and head movement do not provide ground truth for the separation of gaze-driven from gaze-independent head movement. Therefore, we designed a new stimulus and task to elicit both types of movement under controlled conditions in virtual reality (VR). The task involves the presentation of a target the user needs to attain by a gaze shift, followed by the movement of a smaller object from the edge of the target onto its centre using a head gesture. A key goal for our data set was to include head gestures representative of the whole range from the smallest controllable head rotation to the largest comfortable rotation relative to a gaze fixation. We initially tested the range in a pilot study to inform our data collection procedure. We also included the test when we ran the data collection, to validate our data collection design and to provide data to inform gesture design, for instance, with insights on choice of gain factors for head pointing and scale of gestures that are comfortable without disrupting gaze attention. The data collection itself was conducted with 18 participants and yielded over a million timestamped samples of head position, head orientation, eye-in-head orientation, and eye-in-world orientation as input for classification.

With the data collected, we then developed HeadBoost as a novel classifier of Head-Gaze versus Head Gestures. HeadBoost is based on a wide range of features extracted from the data, including spectral, shape-, noise-, time- and correlation-based features, and uses XGBoost [5] for classification. In evaluation, we found HeadBoost to perform with a high F_1-*Score* of 0.89, demonstrating the feasibility of separating Head-Gaze from Head Gestures, and presenting significant improvement over the BimodalGaze heuristic classifier (F_1-*Score*: 0.62). A user-independent model performed even better than a user-dependent model, showing that head movements can be classified with a global approach without requiring individual training. We further show that the HeadBoost model overcomes key limitations of the heuristic baseline in detecting the onset of Head Gestures and Head Gestures performed with slow head movement.

The ability to classify head movement into Head-Gaze and Head Gestures is significant for interaction, as it allows both types of head movement to be leveraged in improved and novel ways. We demonstrate the utility and practical relevance of the classifier with three applications. These show how the classifier supports gestural input with sideways head movement while avoiding Midas Touch [22] by Head-Gaze movements that are performed with similar direction and amplitude. Conversely, we demonstrate Head-Gaze selection of input while avoiding false activation by gestures. Finally, we demonstrate the novel combination of both Head-Gaze and Head Gestures for cursor control, alternating between eye gaze with integral Head-Gaze for gross positioning and head gestures for refinement. In sum, we provide the following contributions:

(1) An introduction to the HCI community of the fundamental differences in head movements driven by gaze (Head-Gaze) and head movements independent of gaze (Head Gestures), to make the case for their treatment as distinct types of input.
(2) A novel stimulus and task design to separate Head-Gaze and Head Gestures for automatic labelling of head movement.
(3) HeadBoost – a novel head-movement classifier for classifying head movements into Head-Gaze or Head Gestures and the results of its performance evaluation. Results show that HeadBoost achieved a F_1-*Score* of 0.89 and outperformed a threshold-based baseline. We further show the benefits of HeadBoost in three VR applications.

2 RELATED WORK

2.1 Classification of Head-Gaze versus Head Gesture

'Head-Gaze' and 'Head Gesture' are widely used notions in the literature, however, not to delineate different types of head movement. 'Head-Gaze' is commonly used in eye tracking literature to describe the approximation of gaze by head pose, to contrast the fine-grained tracking of gaze with an eye tracker that tracks eye-in-head rotation [3]. 'Head Gesture' is a general notion for head movement behaviours associated with specific meaning in human-computer interaction (HCI) literature broadly used for head movements adopted for computer control [26].

In this work, we use the terms to clearly distinguish between head movements that are gaze-driven from head movements that are independent of gaze. This is not inconsistent with the existing use of terms, but provides a basis for a more nuanced consideration of head movement for input. For example, head pointing as an input method implicitly leverages Head-Gaze to move the cursor toward a target when we look at it, but requires an additional Head Gesture to acquire the target [45].

This work is inspired specifically by Sidenmark et al.'s BimodalGaze [46]. The work was not a priori concerned with the classification of head movement, but aimed to use head movement for refinement of gaze input. A naive approach would switch cursor control from gaze to head as soon as a gaze saccade toward a target has been completed. However, if the gaze saccade is supported by head movement, then the head will usually still be in motion after target acquisition, compensated for by VOR eye movement that stabilises vision and rotates the eyes back to a more central position [44]. The continued head movement after gaze acquisition is pre-programmed with the eye-saccade and not under voluntary control until it has been completed. This led Sidenmark et al. to develop an algorithm to filter out 'natural' from 'gestural' head movement to avoid unintended input. This algorithm is based on

thresholds for eye velocity to detect a saccade ($160^\circ/s$), head velocity to detect head movement ($15^\circ/s$), divergence of eye and head trajectory (20°), and delay time between head and eye movement (150ms), derived from eye-head coordination literature [44]. The work is excellent in analysing limitations of the algorithm, leading us to propose that the problem of classifying Head-Gaze versus Head Gesture may be better tackled by machine learning.

There is a wide range of other work on classification of head movements and behaviours. Head movements play a significant role in interaction as a means of expressing intent and eliciting emotions. Previous work has demonstrated the potential of machine learning models for detecting and classifying head gestures in a wide range of application scenarios [4, 15, 40, 58]. For instance, Morency et al. leveraged visual features (e.g., head velocities or eye-gaze estimates) to propose an SVM model to classify head gestures as feedback nods and headshakes during interaction [35]. Similarly, Hachaj and Piekarczyk used PCA-based features to train and propose an artificial neural network that can classify head movements into seven different gestures (e.g., clockwise rotation and head nodding) in a VR environment [15]. Furthermore, researchers in the field of affective computing have explored machine learning models to predict the emotional state of users by leveraging their head movements [58]. Our work is related in using machine learning to model behaviour, but our aim is to fundamentally separate Head-Gaze from Head Gesture.

2.2 Head movement as Input

Head movement has been considered widely for computer control, including pointing, continuous control, spatial selection, and symbolic selection [26]. One major theme is the design of gesture sets, for instance, with recent work proposing nine gestures for select, drag, zoom, scroll, and other commands [54]. A principal problem in gesture design is that gestures need to be robustly distinguishable from any head movement occurring as natural behaviour, often leading to designs that require exaggerated movements [12, 54, 55]. The separation of Head-Gaze from Head Gesture that we propose provides a new route to address this problem and may enable gesture designs of improved usability.

Head movement also lends itself to input in combination with eye gaze, combining the accuracy of head movements with the speed of gaze movements for faster, precise and deliberate interaction [18, 23, 28]. For instance, Kurauchi and Fang adopted Zhai et al.'s MAGIC technique [57] for gaze-assisted head pointing [28], and Sidenmark et al. designed a Look&Cross technique using to pre-select targets in a radial interface that are confirmed by head crossing [47]. Existing work combining the modalities has tended to associate gaze solely with the eyes and a priori separate from head movement. Our work recognises that head movement is integral to gaze, but provides a method to clearly separate those head movements that are part of gaze from those that are usable for complementary input.

3 STUDY DESIGN

Our work is novel in considering the separation of head movement as gaze-driven versus independent of gaze. To develop a classifier, we require data labelled accordingly with ground truth. We approach this with the design of a stimulus and task designed to elicit Head-Gaze and Head Gestures in consecutive phases, which we detail as a novel methodological contribution below.

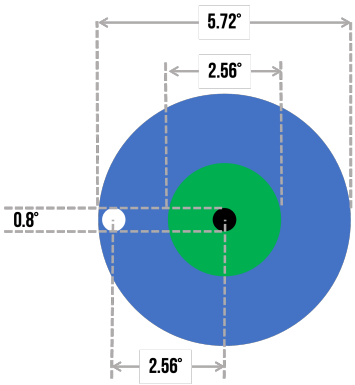


Figure 1: Stimulus (not to scale). Head movement is used to move the white ball into the black hole, over a distance of 2.56° of visual angle in the display space.

A chief concern for the data collection is to include gestural head movements from the very small that may be useful for fine positioning to movements describing larger gestures without disrupting vision. We, therefore, designed a test for head movement range to identify the range from the smallest controllable to the largest comfortable movement. We used this test in a pilot study that we reported as it informed the data collection.

3.1 Stimulus and Task Design

Figure 1 shows the stimulus we designed for data collection, and Figure 2 the task and interaction sequence for eliciting first Head-Gaze and then Head Gestures in two directions relative to the participant's gaze. The interaction sequence follows four phases:

(1) From a neutral forward-looking position, participants performed a gaze shift with associated Head-Gaze movement toward the stimulus.
(2) While their gaze remained on the target, participants performed a Head Gesture to move a ball from the edge of the target onto its centre, using a golfing metaphor.
(3) Participants performed a Head Gesture in the opposite direction, to ensure that we collect both gestural head movement that rotates the head away from the direction of gaze, and toward the direction of gaze.
(4) Participants reset the head position to the neutral, forward-looking position.

With reference to Figure 2, we further elaborate the interaction sequence in the following subsections. The interaction sequence and stimulus design enable automatic labelling of Head-Gaze and Head Gesture data points for the subsequent machine learning classification.

3.1.1 Phase 1: Head-Gaze. Each trial starts from a neutral, forward-looking position. A target appeared in the field of view (FOV) as illustrated in Figure 2a. Participants were tasked to acquire and fixate on the target by gaze, indicated by colour feedback (Figure 2b). To ensure that gaze jitter has no effect on the fixation condition, the

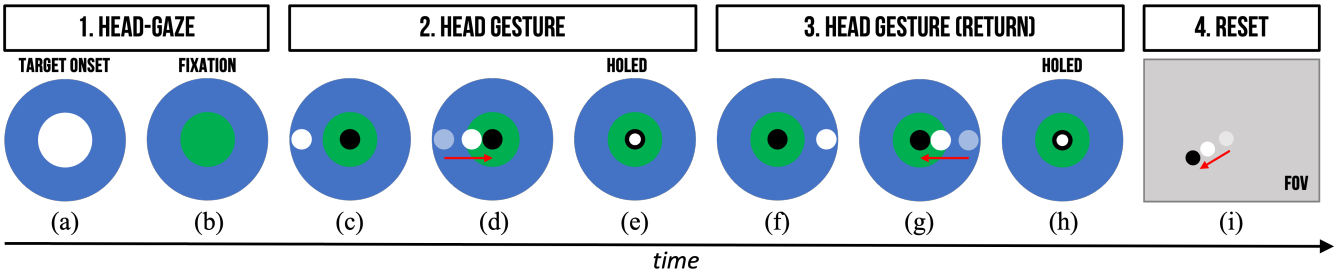


Figure 2: Sequence of events in a trial. Stimulus onset (white inner circle) cues for gaze shift. The inner circle turns green when the participant's gaze hits the target to provide visual feedback. When fixation condition is met, a white ball and a black hole are rendered, which cues the participants to use Head Gesture movement to move the white ball into the black hole while maintaining eye-gaze fixation on the black hole. This is followed by a Head Gesture movement cued in the opposite direction. Finally, the target disappears, and the participant reset to the neutral head position by aligning the white ball to the black ball. The red arrow is used for illustration purposes only.

target is considered fixated on if it is within an area twice the size of the target area. However, the gaze has to be within the target area to start the fixation. While gaze is on target, the Head-Gaze phase continues until head velocity has dropped below 1°/s. We define head velocity as the five-point smoothed velocity [11, 43] of the HMD's Fick-angles [19]. The Head-Gaze phase completes when both conditions are satisfied for 200 milliseconds to ensure that the head is no longer in motion before we prompt gestural movement.

All samples collected in this phase are labelled as 'Head-Gaze'. During this phase, the participant will first move their eyes toward the target, followed by eye and head movement in the same direction during the saccade toward the target. Finally, the eyes and head move in opposite directions when the head is moving to a comfortable position, and the eyes move to remain on target (VOR).

3.1.2 Phase 2: Head Gesture. The second phase of the sequence starts with the appearance of a black hole in the target centre and a white ball at the edge of the target (Figure 2c). The white ball is randomly placed on one of the eight cardinal axes. The participant is then tasked to move the ball towards the hole by turning their head while fixating on the black hole. See Figure 2d, the cued head rotation is illustrated by red arrows.

This golf metaphor requires the participants to monitor the positions of the ball and hole for successful completion. To collect small and precise Head Gesture and larger swiping Head Gesture during this phase, we vary the control-display (CD) gain between the ball on the display and the head-mounted display as the controller:

$$CDGain = \frac{V_{ball}}{V_{hmd}} \quad (1)$$

V_{ball} and V_{hmd} are the velocity of the ball and HMD, respectively. The size of the stimulus is kept constant for data collection. With a larger CD gain, the participant must perform precise pointing head movement to complete the task, while with a smaller CD gain, less precise swiping movement is performed. Note that larger CD gains effectively increases the target width in motor space, such that it can be reached without careful aiming.

Task completion is defined by the following rules and conditions. First, eye-gaze must main fixated on the black hole. Second, the white ball will only move if the performed head rotation is within ±45° of the direction between the white ball and the black hole. This condition enforces that participants pay attention to and perform the head movement in the correct direction. Third, the ball is holed if the position of the ball is within ±0.26° from the centre hole on the display, and the ball's velocity is less than $0.5°/s$ visual angles. We calculate the distance between the ball and the hole via the trapezoidal rule for numerical integration, using the previous and current ball velocities, the previous ball location, and the elapsed time as input. The velocity of the ball is calculated as the head velocity multiplied by the CD gain. When the ball is holed, an animation is shown as visual feedback (Figure 2e). This completes the second phase, during which all data collected are labelled 'Head Gesture'. During this phase, the participant performs head movements and stabilising VOR eye movements to keep the gaze on target.

3.1.3 Phase 3: Head Gesture (Return). The previous phase includes a Head Gesture where the head and gaze align. As such, we include a second Head Gesture phase, where the head moves away from gaze. In this phase, the participant performs the same task as in Phase 2 but in the opposite direction (Figure 2f-h). The head will then return to the starting position of Phase 2. As in Phase 2, all samples are labelled as 'Head Gesture'.

3.1.4 Phase 4: Reset. Finally, participants are guided to reset their head and gaze position to the neural starting position with visual guidance (Figure 2i). This concludes a trial.

3.2 Head Movement Range

We designed a test to assess the range of head movement of interest for gestural input, from smallest controllable to largest comfortable. The test uses the stimulus described above, placed in the centre of the participant's FOV (neutral head position). The independent variable is CD gain, which we manipulate to test head movements that require a different amount of angular rotation to move the ball from the edge of the stimulus into the hole in the centre.

Table 1 shows the 12 levels we used, for each level showing the required amount of head rotation, the CD gain and the effective width of the target in motor space. The total head rotation is the

Table 1: Range of CD gains and their corresponding total head rotation and control target width for assessment of the head movement range usable for gestural input

Total head rotation (°)	50.0	45.0	40.0	32.0	16.0	8.0	4.0	2.0	1.0	0.5	0.25	0.2
CD gain	0.051	0.057	0.064	0.08	0.16	0.32	0.64	1.28	2.56	5.12	10.24	12.8
Control Target Width (°)	10.196	9.123	8.125	6.5	3.25	1.625	0.812	0.406	0.203	0.102	0.051	0.041

range between the minimum and maximum head position angles and is useful for investigating the largest comfortable head rotation. The control target width is the range of head (i.e. control) positions corresponding to the ball (i.e. cursor) locations within the target [1, 2], it represents the range that the head can move while still keeping the ball inside the target area, and is useful for investigating the smallest controllable head rotation.

Participants perform four trials per CD block, for a total of 12 blocks. The Head Gesture directions are randomly sampled out of the eight cardinal directions. Sidenmark and Gellersen [45] investigated the effect of movement orientation on head-eye movement coordination, whereas we randomise movement orientation to identify the general functional movement range. Participants returned to the neutral head position using visual guidance at the end of each trial. After the fourth trial and before the next CD gain block, participants completed a seven-point Single-Ease Questionnaire (SEQ): "Overall, how difficult or easy was the task to complete?", with "1" being the most difficult. Participants started with the middle CD gain to calibrate participants to the SEQ scale. To counterbalance, half of the participants first moved up the CD gain scale before returning to the middle and proceeding to decrease CD gains, while the other half completed the decreasing CD gain scale before proceeding to the increasing CD gains. Then the following sequence starts with a new CD gain and Head Gesture direction.

We conducted a pilot study involving six participants (4M, 2F) to (1) identify a CD gain range suitable for data collection, and (2) to check that our stimulus design works. To manage the complexity of the data collection, we aimed to identify three representative CD gains. We chose the lowest and highest CD gains at which most participants completed the pilot trials with 100% success. The lower bound CD gain was chosen to be 0.08, with a control target width of 6.5°, prompting head movements that are large (32°) and less comfortable to perform. An upper bound CD gain was chosen to be 1.73 with a control target width of 0.3°, the same head-pointing accuracy as reported by Kytö et al. [29], prompting head movements that are small (1.48°) and difficult to control. A third CD gain of 0.32 was chosen to prompt a head movement from the range in which participants had higher accuracy, speed, and SEQ rating.

4 USER STUDY AND DATA COLLECTION

4.1 Participants

We collected data from 18 participants (10F, 8M, 27.7±7.0 years) recruited from our local university. Seven participants had no prior VR experience, 10 reported occasional, 1 reported weekly, and 0 reported daily VR experience. Ten participants had previously used eye tracking as participants in a research setting. Eleven participants wore glasses and seven had good vision without correction.

4.2 Procedure

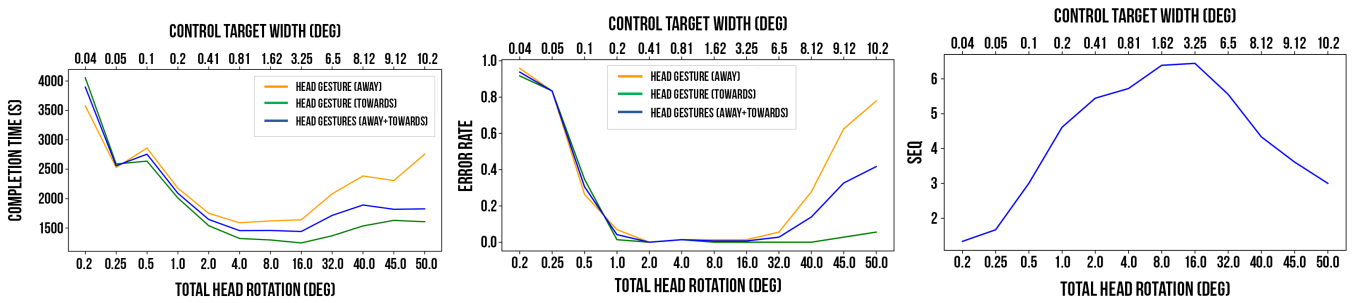
Before starting the session, participants were asked to sit comfortably and given a written overview of the study, a consent form and a basic demographic questionnaire to complete. Participants were briefed on the task, which requires head movement while maintaining gaze on the target, for moving the ball into the hole for each trial. Participants were asked to keep their bodies stationary and only rotate the head to complete the trials as fast and accurately as possible. Further, they were instructed not to touch the HMD or hold it with their hands. Participants were then asked to wear the HMD, adjust the placement of the HMD and the interpupillary distance (IPD), and perform a five-point eye tracking calibration. The details of the headset employed are in the following section.

Participants completed a short practice session to verify that the eye tracking worked for them and that they could comfortably perform the putting task using the three CD gains determined in the pilot study. After completing the practice session, participants performed the Head Movement Range test (subsection 3.2), which took approximately 15 minutes to complete. Once completed, participants were allowed to take a break before continuing with the data collection for our classifier.

We conducted the data collection using three CD gains identified in the pilot study: 0.08, 0.32, 1.73. Participants again completed an eye-tracker calibration task at the start and were reminded to remain still while using the head and focus on the target. Each participant performed three blocks of 25 trials, each block with a different CD gain, in counterbalanced order. In each block, stimuli are placed at a five×five grid of ±30° and were shown in random order. Participants performed one trial per stimulus location, returning to the neutral head position before the next trial began. The direction of the cued Head Gesture was randomised. Participants were allowed to take a break at any time. The data collection took approximately 25 minutes to complete. The study procedure was approved by our institution's ethics board.

4.3 Dataset Description

The study environment and tasks were implemented in Unity version 2020.3.32f1. We collected the eye and head position and directional 3D vectors, relative to the world and head using an HTC Vive Pro Eye (120 Hz). The HMD has a FOV of 100° in the horizontal plane, 110° in the vertical plane and a frame rate of 90Hz. We collected 1192288 time-stamped samples from a total of 81 trials, where each sample is labelled as Head Gesture (Head Gesture (Away): 21%, Head Gesture (Towards): 28%) or Head-Gaze (50%) during data collection. On average per participant, 50.8 ± 5.4% of the sample labels are Head Gesture and 49.2 ± 5.4% are Head-Gaze. The relevant raw data for the classification sections are the head position 3D vector,



(a) Completion time **(b) Error rate** **(c) Single-Ease questionnaire**

Figure 3: Measure to determine the head movement range

head directional 3D vector, eye-in-world directional 3D vector, and eye-in-head directional 3D vector.

4.4 Results on Head Movement Range

From the measure completion time, error rate, and SEQ scores, we estimate a usable range from the smallest controllable head rotation at 0.3° to the largest comfortable rotations at 40° from the gaze position as a general rule of thumb design guide. As every interaction application will have its own intrinsic limitations, we determined this range quantitatively from Figure 3, without applying threshold-based cutoffs.

Task completion time increases as the CD gain is too sensitive (low control target width) or too large, causing the target to exit the FOV (Figure 3a). The error rate increases as the control target width decreases, or as the total head rotation is exaggerated (Figure 3b). In the latter case, Head Gesture (Away) is more affected because it causes the target to exit the FOV. Participants found the task easy to complete when the head movement is moderate (Figure 3c). We observed that the error rate becomes impractical and rises steeply after $\sim 40°$ of total head rotation, and in line with previous research, below approximately 0.3° of control target width.

5 CLASSIFICATION OF HEAD MOVEMENTS INTO HEAD-GAZE AND HEAD GESTURE

This section describes the pre-processing, feature extraction, and classification methods employed to classify head movements into Head-Gaze and Head Gesture.

5.1 Preprocessing

We pre-processed the raw data before extracting features for classification according to best practices [7, 8]. First, we removed data samples with the calculated velocity of $> 800°/s$ according to physiological limits in eye rotation. Second, we interpolated the data samples to a constant sampling rate of 125Hz. Next, we transformed the 3D directional gaze vectors into 2D Fick angles via the Fick-gimbal, where the eye (or head) position is characterised by a rotation about the vertical axis (Azimuth, or Az angle) followed by a second rotation about the nested horizontal axis (Polar, or Pol angle) [19]. Lastly, we calculated the velocities and accelerations of the eye-in-world and eye-in-head vectors, head angles, and head positions using cubic spline derivatives.

5.2 Feature Extraction

We employed common features for eye- and head-based classification [9, 21, 31, 37, 41, 48, 51, 56]. Table 2 summarises these features, which could be categorised into shape-, noise-, spectral-, correlation-, and timing-based features, interested readers can find their precise implementations in the original papers cited in Table 2. Previous work using gaze-based features for classification [25, 27] suggests that window length significantly affects the performance of the classification model. Therefore, we experimented with five different window lengths to generate the features for the classification: 128ms, 256ms, 312ms, 384ms, 512ms. We chose a window length of 512ms, as it gave the highest classification performance using 5-fold cross-validation, reaching a plateau.

5.2.1 Feature Selection. Research suggests that training a machine learning model on a large feature set results in high computational costs and potentially leads to overfitting the model [10]. Therefore, various feature selection strategies [6, 33, 42] are deployed that select a small subset of relevant features by removing redundant and noisy features from the original feature set. Consistent with this observation, we used the correlation-based feature selection method [16] to select relevant features for the classification task. For this purpose, we first calculated the correlation distance between each feature. We then used the calculated correlation distances to cluster features using hierarchical clustering [36]. Lastly, we selected a single feature from each feature cluster to obtain 81 relevant features to classify each head movement.

5.3 Classification

After feature engineering and selection, we combined the feature vector in a machine learning model that could classify the type of head movement. We modelled this task as a binary classification problem. For each head movement trial, the selected features were fed into a classifier to predict the type of head movements as either Head-Gaze or Head Gesture. To train a machine learning model, we experimented with eXtreme Gradient Boosted (XGBoost), SVM and Random Forest models by training them on the collected dataset. We selected XGBoost as the final model for the classification task, as it gave the highest performance score across the testing folds. XGBoost is an optimised distributed implementation of the gradient-boosted decision tree algorithm specifically designed to be highly accurate, fast, and flexible [5]. The XGBoost model primarily solves

Table 2: HeadBoost features.

Category	Features
Shape-based	Slope, Range, Mean Velocity, Peak Velocity, Mean Acceleration, Peak Acceleration, Integral, Energy, Wavelength [51], Spatial features in the positional signal (P_D, P_{CD}, P_{PD}, P_R) defined by Larsson et al. [31]
Noise-based	Dispersion [41], Standard Deviation, RMS, BCEA [21, 48, 56], RMS-diff, BCEA-diff, Mean-diff, Median-diff [37, 56], Rayleightest [31, 56]
Spectral	Rolloff, Centroid, Entropy [9], Flatness
Correlation-based	Correlation
Timing-based	Time since last saccade (200 $^\circ/s$)

the classification problem by combining an ensemble of estimates from a set of simpler and weaker tree models. We built the XGBoost classifier with 20 trees on the whole dataset of 18 participants.

As in previous work, we evaluated HeadBoost using the F_1-Score and Area Under the Receiver Operating Characteristics Curve metrics (AUC) [14]. The F_1-Score [13] is the weighted average of Precision and Recall and calculated as $2 \times \frac{Precision \times Recall}{Precision + Recall}$, where $Precision = \frac{TP}{TP+FP}$ and $Recall = \frac{TP}{TP+FN}$, and TP, FP, and FN represent the number of true positives, false positives, and false negatives, respectively. The Receiver Operating Characteristics (ROC) curve represents the relationship between the true positive rate and the false positive rate at different classification thresholds. AUC measures the area under the ROC curve and represents the ability of a classifier to distinguish between classes [17]. F_1-Score and AUC scores range from 0 to 1, where 1 represents the highest and 0 represents the lowest possible performance values the classifier could attain. A score of 0.5 is typical of random guessing.

To capture the temporal history of participant behaviour, for each observation of the head movement, we included the feature vector from the last T ms, sampled at every s Hz. We experimented with different combinations of temporal history windows (T = 256ms, 512ms, 640ms and 1024ms) and sampling rates (i.e., s = 12.5Hz, 6.25Hz, 3.13Hz, 1.56Hz and 0.98Hz) for building our classifier. Finally, we selected a temporal history window (T) of 1024 ms with a sampling rate of 6.25Hz because the model gave the highest classification performance on the testing folds for this combination and reached a performance plateau for T>1024ms and s>6.25Hz.

We evaluated our classification approach in two ways. We first performed a *user-dependent* classification by training and evaluating the classifier on the data from the same participant but from a different head movement trial. We build the user-dependent classifiers using five-fold cross-validation. For example, if a participant conducted 75 head movements trials, we trained the classifier five times, each time training on the data of four folds–containing 60 trials each—and evaluated it on the data of the remaining fold, which includes 15 data trials. The reported results (see Table 3) are averaged across the five folds and the total number of participants.

To avoid overfitting the classifier to the behaviour of a particular participant and propose a generic classifier, we further conducted a *user-independent* evaluation. We evaluated the classifier by training it 17 times using leave-one-participant-out cross-validation. For this purpose, each time we trained the classifier on the data of 16 participants (one participant was removed from the training set due to noise but was used in the evaluation set) and then evaluated it on the head movement trials of the last participant. The reported results in Table 3 are averaged by the total number of participants.

Table 3: Performance of the baseline [46] and our proposed approach (HeadBoost) to classify head movements.

Evaluation Measures	Baseline Approach [46]	Proposed Models	
		user-dependent	user-independent
F_1-Score	.62 ± .06	.87 ± .05	.89 ± .06
AUC	.66 ± .03	.94 ± .02	.96 ± .02
Precision	.71 ± .09	.86 ± .05	.88 ±.01
Recall	.58 ± .11	.88 ± .04	.91 ± .04

6 EVALUATION OF THE CLASSIFIER

In this section, we first report the performance of the user-dependent and user-independent models trained to classify head movements. Then we compare our proposed approach with the threshold-based approach proposed by Sidenmark et al. [46]. Lastly, we explore the features that significantly impact the performance of the user-independent classifier for head movement classification.

6.1 Model Performance

We computed the AUC, F_1-Score, recall and precision to evaluate the performance of the built classifiers (see Table 3). Our results suggest that it is generally feasible to classify head movements into Head-Gaze and Head Gesture (F_1-Score = 0.89). Furthermore, we observed that the user-independent classifier could more accurately classify head movement compared to the user-dependent classifier (see Table 3). This is because, in contrast with the user-dependent classifier, the user-independent classifier is trained on a much larger volume of data. Thus, it learns patterns from a wide range of head movements depicted by various participants, consequently classifying new head movements more precisely.

Further, we observed that although the generic user-independent model has a high classification performance (F_1-Score = 0.89), it fails to predict the correct label of head movement for a few trials. An exploratory analysis of the data suggests that one of the reasons for misclassification was when participants made an unintentional head movement in the trials. In some cases, the user-independent

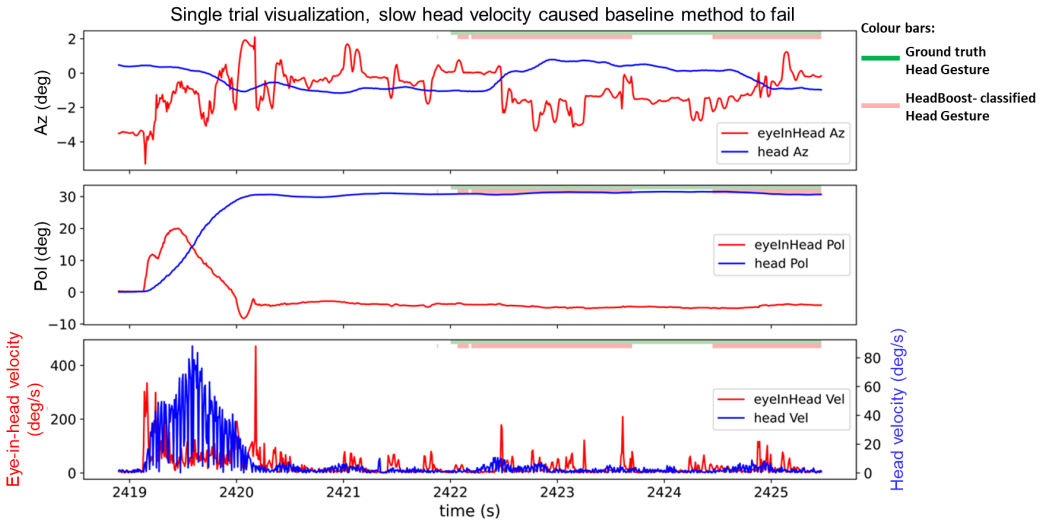


Figure 4: Slow head movement (< 15°/s) caused the baseline method to miss the Head Gesture. Single-trial visualization of the eye and head rotation and velocities. At the top of the graphs, the green bar shows the ground truth label for Head Gesture; the pink bar shows the predicted Head Gesture by the HeadBoost model.

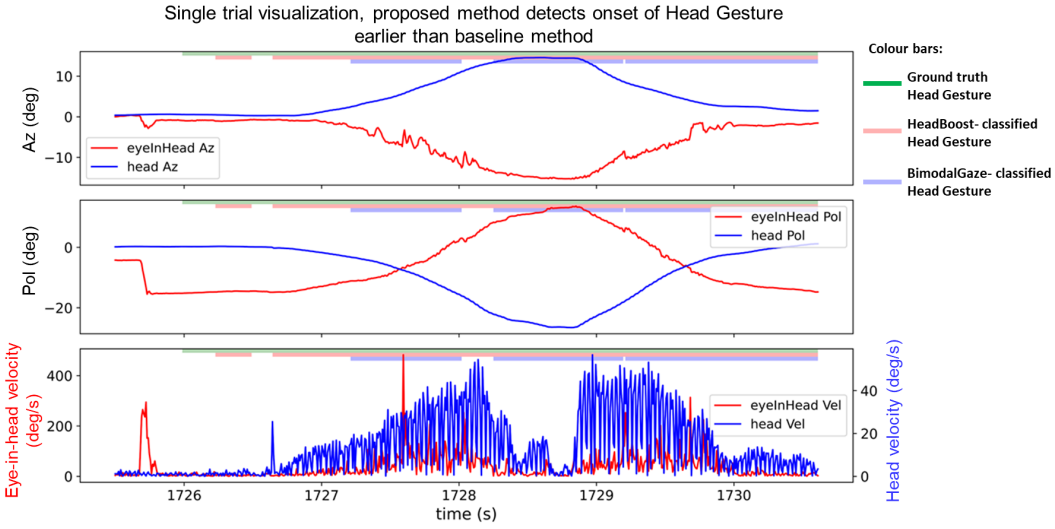


Figure 5: The HeadBoost model detects the onset of Head Gesture earlier than the baseline method. Single-trial visualization of the eye and head rotation and velocities. At the top of the graphs, the green bar shows the ground truth label for Head Gesture; the pink bar shows the predicted Head Gesture by the user-independent XGBoost classifier. The purple bar shows the prediction by the baseline method. Az and Pol refer to the Fick angles that characterize the gaze direction (see subsection 5.1)

classifier would incorrectly predict the head movement as Head Gesture, even though a natural Head-Gaze movement has occurred. In other cases, the Head Gesture is predicted correctly, but the ground truth is corrupted by the unintentional Head Gesture during data collection when they were instructed to perform Head-Gaze. This behaviour of the classifier increases the false-positive rate of the model. However, the effect is difficult to quantify due to the subjective nature of interpreting the ground truth labels and human behaviour, which we will elaborate in the discussion.

6.2 Comparison with Baseline Method

To demonstrate the effectiveness of HeadBoost, we compared the performance of our approach with the threshold-based approach of Sidenmark et al. [46]'s *BimodalGaze*. The F_1-Score is increased from 0.62 to 0.89, indicating a substantial improvement in overcoming limitations in classification sensitivity and speed. This increase in performance could be explained by the fact that our approach learns the pattern of head movement using additional head-, gaze-, and time-based features to perform the classification task. However,

Sidenmark et al. use fixed thresholds for various features to classify head movement. Consequently, their method may break if the feature values deviate from the fixed thresholds. For example, we observed that whenever participants' head velocity was constantly below a fixed threshold of $15°/s$, the baseline method incorrectly classified the head movement as Head-Gaze. However, as shown in Figure 4 this error is resolved in our proposed approach as it uses a machine learning model to learn the pattern of head movement rather than relying on a fixed head velocity value.

We further observed that our proposed method was able to predict the onset of Head Gesture much earlier than the threshold-based method [46] (avg. 119ms earlier for all trials). For example, Figure 5 shows that the proposed method detected the onset of Head Gesture between the time stamp of 1726s and 1727s (depicted by the horizontal pink bar). However, the baseline method detected Head Gesture at a later timestamp—between 1727s and 1728s (illustrated through the horizontal purple bar).

6.3 Feature Importance

We used the SHapley Additive exPlanations (SHAP) algorithm [32] to explore the importance of each feature in classifying head movement. Figure 6 illustrates the feature importance plot, where the features are ordered in decreasing importance from top to bottom. As seen in Figure 6, no single feature is solely responsible for classification. Rather, a combination of head-based and timing-based features contributed to predicting the types of head movement.

We observed that the historical value of *Rayleightest*, which captures whether the head movement was uniform 1024 ms ago, contributes the most to correctly predicting the type of head movement. This is because, for Head-Gaze observations, a uniformly distributed sample-to-sample head direction pattern is often observed, resulting in a high value of this feature. On the other hand, for Head Gesture observations, the sample-to-sample head direction distribution is non-uniform, resulting in low feature values.

Furthermore, we observed that some shape-based head features (e.g., head wavelength and P_D value that shows the relationship between successive principal components of head movement) are important for predicting the head movement types. Similarly, some noise-based head features, such as the *RMS-Diff* feature, which reflects the difference in the root mean square value of head movement between successive feature-generation windows, also got high importance for the head movement classification task.

Lastly, we observed that the timing-based feature, total time since the last saccade, played a signification role in distinguishing head movements. This observation could be explained by the fact that Head Gestures generally occur after the gaze shift is completed and a target is acquired. Therefore, the time since the last saccade would be much longer for Head Gestures than for Head-Gaze.

Further, we observed that HeadBoost correctly detected Head Gesture very early for some trials, as shown in Figure 5. To identify the features that highly impacted the decision of HeadBoost for early detection, we performed a SHAP analysis using the first detected Head Gesture period illustrated in Figure 5. We observed that for early Head Gesture detection, *Head (Az-Pol combined) wavelength* was the most important feature, followed by the historical

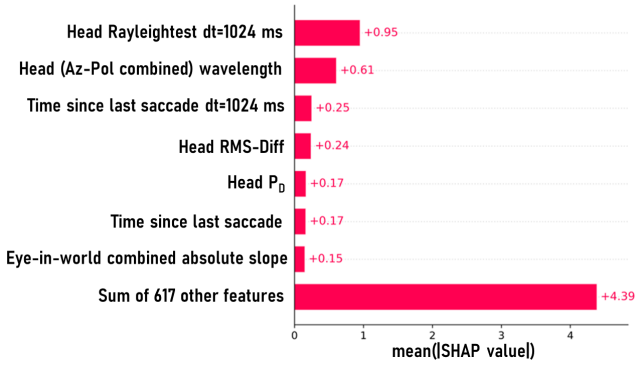


Figure 6: SHAP value feature importance of a user-independent classifier

values of *Rayleightest*, *wavelength*, and *eye-in-world combined absolute slope* features. This observation suggests that the early detection of the Head Gesture is impacted by the temporal evolution of the features and the current head and eye movement.

7 APPLICATIONS

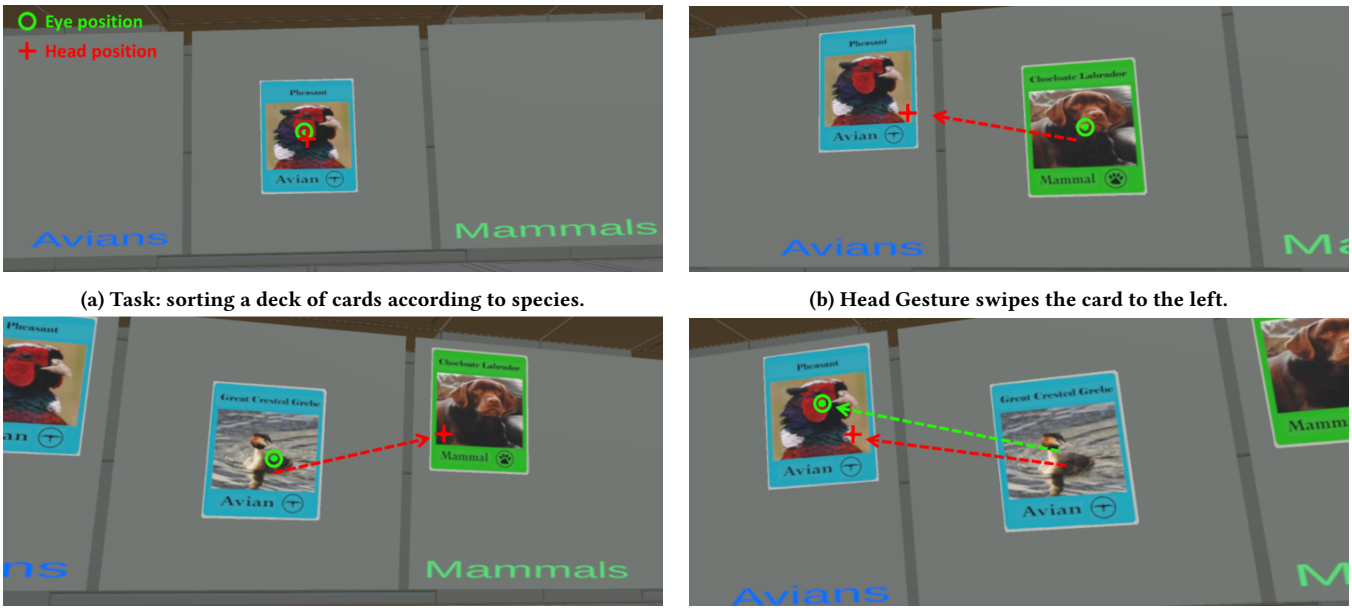
We developed three applications to demonstrate the utility of our classifier. The applications and their features are summarised in Table 4. In Application 1, we demonstrate that, in contrary to the complex and repetitive gestures used in previous works, simple head gestures can be used as input commands without accidental activation from Head-Gaze. As shown in Figure 7, the user performs simple swiping gestures to the left or right to sort cards into the appropriate bin. Later, the user performs a gaze shift in reaction to external stimuli. HeadBoost then recognises the movements as part of the Head-Gaze and filters out the movement. Therefore, no swipe interaction is performed.

In Application 2, we build on the separation of Head-Gaze and Head Gesture and demonstrate that multiple, separate head-based interactive systems can coexist without accidentally triggering each other. As shown in Figure 8, the application consists of a virtual workspace where the user is surrounded by multiple virtual displays in a horizontal layout [34]. The current display is mapped to the head direction. If the user performs Head-Gaze far enough in either horizontal direction, the system will change the current attended display. Users can interact with content within each display using their head, in this case, the same swiping interaction as in Application 1. As multiple interactive systems depend on the head, it is important to separate horizontal swiping gestures within a window from Head-Gaze performed to switch displays. We show that Head-Gaze can afford attention-based workspace switching while allowing Head Gesture as control input within the attended workspace without accidentally switching displays.

In Application 3, we demonstrate how the classification of Head-Gaze and Head Gesture affords seamless mode-switching from gaze pointing to head pointing and that HeadBoost is relevant for fine-grained interaction beyond simple swipe-style applications. As shown in Figure 9, this application is an adaptation of the *Bimodal-Gaze* pointing technique by Sidenmark et al. [46]. The pointer is primarily controlled by gaze, but when the user has to refine the

Table 4: The different affordances of Head-Gaze and Head Gesture in applications and the benefit provided by HeadBoost.

Application	Head-Gaze	Head Gesture	Benefit
1	Filtered out and ignored	Swipe interaction to sort cards	Avoid accidental activation by Head-Gaze
2	Switch between multiple windows	Swipe interaction to sort cards in the attended window	Separate head-based applications can coexist
3	Gaze pointing	Head movement refines cursor position	Head-Gaze and Head Gesture can be used in the same application, depending on user needs



(a) Task: sorting a deck of cards according to species.

(b) Head Gesture swipes the card to the left.

(c) Head Gesture swipes the card to the right.

(d) Head-Gaze to look at an already sorted card to the left without triggering a swipe action.

Figure 7: Application 1: The participant sorts a stack of cards of animals of avians and mammals using Head Gesture. When a Head-Gaze is performed to look at an already sorted card to the left (or right), it does not trigger a swipe action. Head-Gaze is suppressed to allow simple head gestures in interactions without false-active activation. The green circle is the eye position. The red cross-hair is the head position. These are for illustration purposes only.

cursor position when selecting small targets or due to tracking difficulties [29], the head can perform small gestural movements to move the cursor to the correct position. With HeadBoost, head refinement is activated when the classifier identifies a Head Gesture. A switch back to gaze-pointing occurs when the user performs a saccade or a head movement triggered by gaze. This application shows how HeadBoost's classification of Head-Gaze and Head Gesture can be leveraged in the same interaction technique to switch interaction modes depending on current needs.

8 DISCUSSION

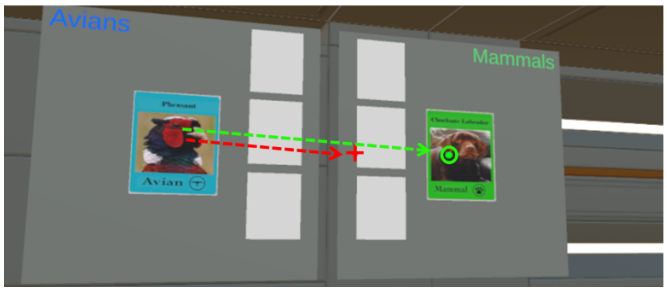
8.1 Head-Gaze versus Head Gesture

We introduced the problem of classifying Head-Gaze and Head Gestures, motivated by fundamental differences in the nature of the underlying head movements, and their affordances for interaction. In existing work, head movement is treated as a uniform input type, whereas we argue for their treatment as distinct modes.

Previous research has demonstrated that head input can improve the user experience by making input more seamless and effortless [15, 38, 54]. However, these may use unusual gestures, repetitions, or dwell to avoid false activation by Head-Gaze. HeadBoost shows the potential for simpler and more natural gestures to be used as inputs with low false activation. In our applications, we demonstrate that different types of input can be assigned to Head-Gaze and Head Gesture to capitalise on their respective advantages, e.g., naturally adjusting the interface according to Head-Gaze as input focus changes or using Head Gesture for fine control of UI elements.

As eye tracking enabled AR/VR devices to evolve to provide low-friction, always-on assistance [24], we anticipate the ability to distinguish Head-Gaze and Head Gesture to become more relevant to HCI. HeadBoost supports this notion by being device-independent,

(a) Task: interacting with cards via Head Gesture, switching windows with Head-Gaze. The current workspace is the Avians window (centre).



(b) User wishes to switch to the Mammals window and performs a Head-Gaze to it without accidentally triggering a card swipe.

(c) Head-Gaze to the Mammals window moves it to the centre.

(d) User performs Head Gesture to view the next card without accidentally switching to the Avians window.

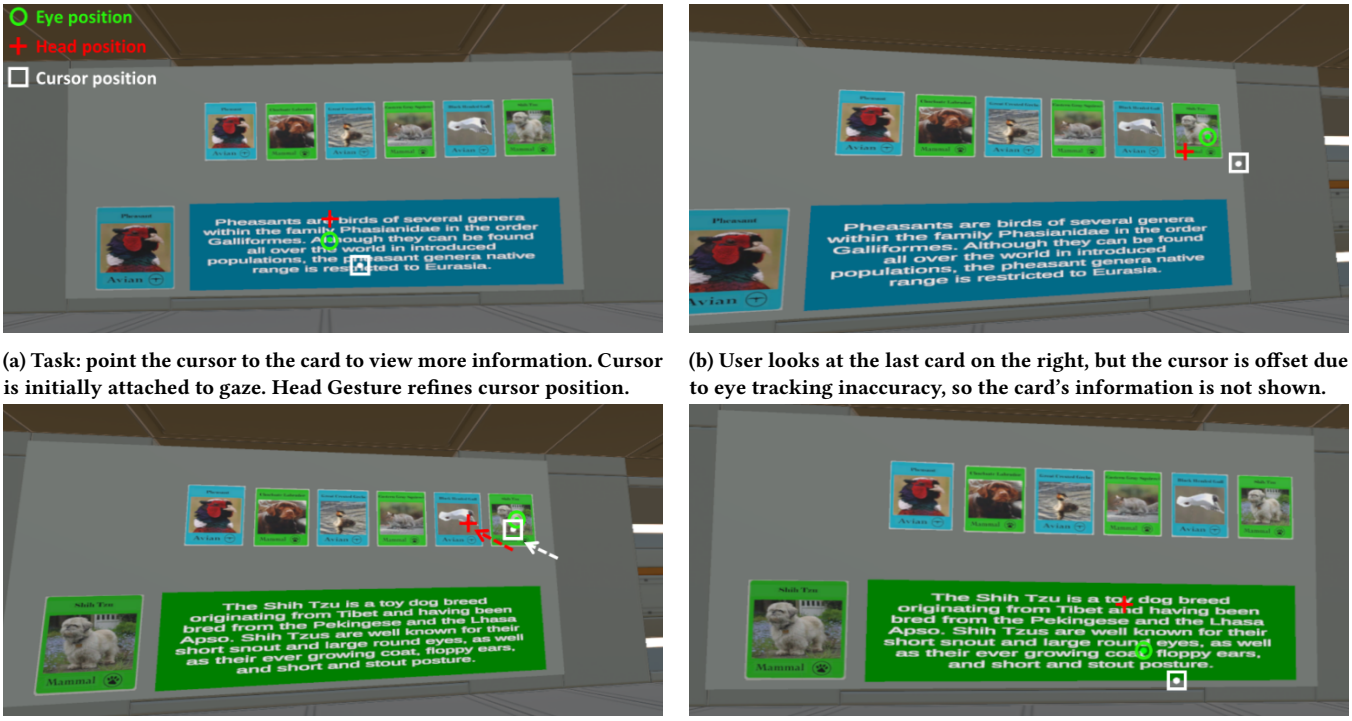
Figure 8: Application 2: Two windows are presented to the user, the central window is the current workspace. Head-Gaze is used to switch the attended window to the centre, and Head Gesture is used to view the cards inside the current workspace without accidentally switching to another window. The green circle is the eye position. The red cross-hair is the head position. These are for illustration purposes only and are not visible to the user.

enabling easy transfer between devices by only requiring the final output of the eye and head trackers without device-specific parameters such as eye images. As modern eye trackers provide a sample-wise classification of fixations, saccades, and other gaze events, we anticipate it would be useful to also provide the classification of Head-Gaze and Head Gesture. This classification would be helpful to enable activity recognition and behavioural studies beyond the desktop.

8.2 HeadBoost

We proposed classifying head movements into Head-Gaze versus Head Gesture to facilitate their treatment as different input types. We developed and evaluated HeadBoost, an XGBoost classifier on the collected dataset. Our proposed method achieved a high classification performance for the user-independent model (F_1-Score = 0.89), promising a global approach where a new head movement can be classified without individual training. Furthermore, we observed that our proposed method outperforms the threshold-based baseline method [46] by an F_1-Score of 0.27, and detects the onset of Head Gesture 119 ms earlier than the baseline method. Our approach boosts performance because it aims to learn the pattern of eye-head movement to make predictions instead of relying on a fixed threshold of features. Furthermore, AR/VR developers may find our work informative as a guide for the future development of head gesture-controlled interfaces. We found a head movement range between the smallest controllable (0.3°) and the largest comfortable (40°), where participants had a lower error rate, shorter task completion time, and a higher SEQ score.

Another challenge for machine learning classification problems is the data's ground truth labels. We introduced a task that automatically separates the Head-Gaze and Head Gesture phases instead of relying on costly and subjective manual labour. In previous work, gestural head movements can be clearly separated based on their repetitive and unusual nature [15, 38, 54], and the classification output is often not available until the gesture motion is completed or has been happening for a period. Our work aims to separate simple gestures from Head-Gaze in real-time. Therefore, it is more difficult to distinguish whether a movement is a Head Gesture, and manual labelling will likely involve more subjectivity. Simply swiping an object sideways compounds the Head-Gaze and Head Gesture movements, as the participant has to monitor the object's position with their foveal vision. The golf metaphor, in contrast, requires the participant to fixate on the centre hole and monitor the ball position using their periphery vision, clearly separate Head-Gaze and Head Gesture. It also allows us to investigate the usable head movement range, as participants must accurately control the ball position while fixating on the target. Furthermore, we manipulated the CD gains to draw out Head Gesture of different amplitudes and velocities to create a data set with more variance.



(a) Task: point the cursor to the card to view more information. Cursor is initially attached to gaze. Head Gesture refines cursor position.

(b) User looks at the last card on the right, but the cursor is offset due to eye tracking inaccuracy, so the card's information is not shown.

(c) Head Gesture attaches cursor to head movement for fine-grained cursor refinement to select and view the card's information.

(d) Head-Gaze to look at the information at the bottom switches cursor control to gaze pointing.

Figure 9: Application 3: Users point at cards to view more information. Initially, the cursor is attached to gaze. Due to eye tracker inaccuracy, an offset prevents the selection of the intended card. Head Gesture is used to switch to head pointing for fine-grained cursor refinement. Cursor control is switched back to gaze pointing when Head-Gaze is detected. The green circle is the eye position. The red cross-hair is the head position. The white square is the cursor position. These are for illustration purposes only and are not visible to the user.

8.3 Limitations and Future Work

We acknowledge five limitations of our work. First, participants in our study were asked to keep the body stationary and only rotate the head to complete the trials. Hence, our result is limited to a sitting position, with targets in the ±30 range in both vertical and horizontal directions of the FOV. Future work would investigate standing users with free torso movements. During motion, the VOR is involved in stabilising gaze on the object of interest. Even though the head itself may not be rotating, we may observe eye-head movement patterns, which may be confused as Head Gesture due to the translation of the torso in the world. The dynamics of in-the-wild use are more complex. Therefore, classification during movement would present another challenge for future work.

Second, we employed a VR headset with base station tracking. In-the-wild systems may not have as accurate tracking systems as in a lab environment. We should investigate whether this affects the use of the world-coordinate related features, which were found to be important for HeadBoost. Future research should aim to collect and train on in-the-wild data and utilise sensor fusion methods and scene recognition for anywhere position inference without a defined tracking area.

Third, while the online classifier showed promising application, it is limited to 30Hz due to having to calculate a large number of features. We plan to improve the classification process by further dimensionality reduction. A reduced feature set without loss of accuracy could lead to a lighter-weight classifier that may run on untethered headsets.

Fourth, the generalisability of HeadBoost could be further evaluated with user studies in natural applications. Although carried out in a controlled experiment, the tasks on which HeadBoost was trained capture any intended head stroke about an intended target. However, every application will have its own influences and needs for optimisation. Our three applications serve as proof of concept to show how the classifier could benefit real-life scenarios. However, we have yet to formally evaluate them with user studies, which we aim to conduct in future work.

Further, in future work, the effect of unintentional head movement could be investigated. HeadBoost separates different types of head movement as gaze-driven or gaze-independent. Either type can involve intention depending on context. We do not classify whether a head movement is performed with an intention. As such, classified Head Gestures can be "intentional" for interacting with the interface or "unintentional" while still being unrelated to gaze

(e.g., looking up while thinking). We observed that if the participant performed an unintentional Head Gesture during the period they were instructed to perform Head-Gaze, the ground truth label would conflict with the actual movement type, which caused misclassification. The extent of this effect is difficult to quantify because it is challenging to collect and label movements as intentional or unintentional. Any manual labelling of the data set would likely be time-consuming and pose another challenge of evaluating human judgement of intentions. Our tasks aim to collect intentional Head Gesture, although random, unintentional head movement that is independent of gaze can still occur during the Head-Gaze phase, the portion of these data samples is small compared to correctly performed movement. As supported by the F_1 and AUC scores, the classifier likely captured the patterns of Head-Gaze and Head Gesture. Hence, we could further characterise the effect of unintentional movement on the classifier through a user study.

9 CONCLUSION

In this work, we demonstrate that it is feasible to distinguish two fundamentally different head movements – Head-Gaze, which supports the eyes in gaze shift, and the gaze-independent Head Gesture, and that they can be treated as different inputs. We proposed HeadBoost, a user-independent XGBoost classifier based on shape, noise, spectral, correlation, and temporal features that achieved a F_1-Score of 0.89, significantly above the baseline (F_1-Score = 0.62). We demonstrate its online utility and practicality in three applications: the classifier supports gestural input while avoiding Midas Touch by Head-Gaze; Head-Gaze selection of input focus while avoiding false activation by gestures; and switching of cursor control between eye gaze (with integral Head-Gaze) and Head Gesture for refinement. Furthermore, we report the finding that the smallest controllable head rotation is about 0.3°, and the largest comfortable rotation is approximately 40°. The classification of Head-Gaze and Head Gesture enables designers to create a more seamless and natural way of interaction using simple head movements while avoiding false activation.

ACKNOWLEDGMENTS

We would like to thank Dominic Potts for his work on the applications. This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant No. 101021229 GEMINI: Gaze and Eye Movement in Interaction).

REFERENCES

[1] Lynn Y. Arnaut and Joel S. Greenstein. 1987. An Evaluation of Display/Control Gain. In *Proceedings of the Human Factors Society Annual Meeting*, Vol. 31. SAGE Publications, Los Angeles, CA, USA, 437–441. https://doi.org/10.1177/154193128703100412

[2] Lynn Y. Arnaut and Joel S. Greenstein. 1990. Is Display/Control Gain a Useful Metric for Optimizing an Interface? *Human Factors* 32, 6 (1990), 651–663. https://doi.org/10.1177/001872089003200604

[3] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of Eye-Gaze over Head-Gaze-Based Selection in Virtual and Augmented Reality under Varying Field of Views. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 1, 9 pages. https://doi.org/10.1145/3206343.3206349

[4] Riccardo Bovo, Daniele Giunchi, Ludwig Sidenmark, Hans Gellersen, Enrico Costanza, and Thomas Heinis. 2022. Real-Time Head-Based Deep-Learning Model for Gaze Probability Regions in Collaborative VR. In *2022 Symposium on Eye Tracking Research and Applications* (Seattle, WA, USA) *(ETRA '22)*. Association for Computing Machinery, New York, NY, USA, Article 6, 8 pages. https://doi.org/10.1145/3517031.3529642

[5] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (San Francisco, California, USA) *(KDD '16)*. ACM, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

[6] Xue-wen Chen and Jong Cheol Jeong. 2007. Enhanced Recursive Feature Elimination. In *Proceedings of the Sixth International Conference on Machine Learning and Applications (ICMLA '07)*. IEEE Computer Society, USA, 429–435. https://doi.org/10.1109/ICMLA.2007.44

[7] Antoine Coutrot, Janet H. Hsiao, and Antoni B. Chan. 2018. Scanpath modeling and classification with hidden Markov models. *Behavior Research Methods* 50, 1 (01 Feb 2018), 362–379. https://doi.org/10.3758/s13428-017-0876-8

[8] Brendan David-John, Candace Peacock, Ting Zhang, T. Scott Murdison, Hrvoje Benko, and Tanya R. Jonker. 2021. Towards Gaze-Based Prediction of the Intent to Interact in Virtual Reality. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Short Papers)*. Association for Computing Machinery, New York, NY, USA, Article 2, 7 pages. https://doi.org/10.1145/3448018.3458008

[9] Michael Dietz, Daniel Schork, Ionut Damian, Anika Steinert, Marten Haesner, and Elisabeth André. 2017. Automatic Detection of Visual Search for the Elderly using Eye and Head Tracking Data. *KI - Künstliche Intelligenz* 31, 4 (01 Nov 2017), 339–348. https://doi.org/10.1007/s13218-017-0502-z

[10] R.P.W. Duin. 2002. The combining classifier: to train or not to train?. In *2002 International Conference on Pattern Recognition*, Vol. 2. IEEE, 765–770 vol.2. https://doi.org/10.1109/ICPR.2002.1048415

[11] Ralf Engbert, Lars OM Rothkegel, Daniel Backhaus, and Hans Arne Trukenbrod. 2016. Evaluation of velocity-based saccade detection in the SMI-ETG 2W system. *Technical report, Allgemeine und Biologische Psychologie, Uni-versität Potsdam, March* (2016).

[12] Wenxin Feng, Jiangnan Zou, Andrew Kurauchi, Carlos H Morimoto, and Margrit Betke. 2021. HGaze Typing: Head-Gesture Assisted Gaze Typing. In *ACM Symposium on Eye Tracking Research and Applications* (Virtual Event, Germany) *(ETRA '21 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 11, 11 pages. https://doi.org/10.1145/3448017.3457379

[13] Cyril Goutte and Eric Gaussier. 2005. A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation. In *Advances in Information Retrieval*, David E. Losada and Juan M. Fernández-Luna (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 345–359. https://doi.org/10.1007/978-3-540-31865-1_25

[14] Qiong Gu, Li Zhu, and Zhihua Cai. 2009. Evaluation Measures of the Classification Performance of Imbalanced Data Sets. In *Computational Intelligence and Intelligent Systems*, Zhihua Cai, Zhenhua Li, Zhuo Kang, and Yong Liu (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 461–471. https://doi.org/10.1007/978-3-642-04962-0_53

[15] Tomasz Hachaj and Marcin Piekarczyk. 2019. Evaluation of Pattern Recognition Methods for Head Gesture-Based Interface of a Virtual Reality Helmet Equipped with a Single IMU Sensor. *Sensors* 19, 24 (2019), 19 pages. https://doi.org/10.3390/s19245408

[16] Mark A Hall. 1999. *Correlation-based feature selection for machine learning*. Ph. D. Dissertation. The University of Waikato.

[17] J A Hanley and B J McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143, 1 (1982), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

[18] John Paulin Hansen, Vijay Rajanna, I. Scott MacKenzie, and Per Bækgaard. 2018. A Fitts' Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-Mounted Display. In *Proceedings of the Workshop on Communication by Gaze Interaction* (Warsaw, Poland) *(COGAIN '18)*. Association for Computing Machinery, New York, NY, USA, Article 7, 5 pages. https://doi.org/10.1145/3206343.3206344

[19] Thomas Haslwanter. 1995. Mathematics of three-dimensional eye rotations. *Vision Research* 35, 12 (1995), 1727–1739. https://doi.org/10.1016/0042-6989(94)00257-M

[20] Dirk Heylen. 2006. Head gestures, gaze and the principles of conversational structure. *International Journal of Humanoid Robotics* 03, 03 (2006), 241–267. https://doi.org/10.1142/S0219843606000746

[21] Kenneth Holmqvist, Marcus Nyström, and Fiona Mulvey. 2012. Eye Tracker Data Quality: What It is and How to Measure It. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 45–52. https://doi.org/10.1145/2168556.2168563

[22] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (apr 1991), 152–169. https://doi.org/10.1145/123078.128728

[23] Shahram Jalaliniya, Diako Mardanbegi, and Thomas Pederson. 2015. MAGIC Pointing for Eyewear Computers. In *Proceedings of the 2015 ACM International Symposium on Wearable Computers* (Osaka, Japan) *(ISWC '15)*. Association for

Computing Machinery, New York, NY, USA, 155–158. https://doi.org/10.1145/2802083.2802094

[24] Tanya R Jonker, Ruta Desai, Kevin Carlberg, James Hillis, Sean Keller, and Hrvoje Benko. 2020. The Role of AI in Mixed and Augmented Reality Interactions. In *CHI2020 ai4hci Workshop Proceedings*. ACM.

[25] Peter Kiefer, Ioannis Giannopoulos, and Martin Raubal. 2013. Using Eye Movements to Recognize Activities on Cartographic Maps. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (Orlando, Florida) *(SIGSPATIAL'13)*. Association for Computing Machinery, New York, NY, USA, 488–491. https://doi.org/10.1145/2525314.2525467

[26] R. Kjeldsen. 2001. Head gestures for computer control. In *Proceedings IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*. IEEE, 61–67. https://doi.org/10.1109/RATFG.2001.938911

[27] Kai Kunze, Yuzuko Utsumi, Yuki Shiga, Koichi Kise, and Andreas Bulling. 2013. I Know What You Are Reading: Recognition of Document Types Using Mobile Eye Tracking. In *Proceedings of the 2013 International Symposium on Wearable Computers* (Zurich, Switzerland) *(ISWC '13)*. Association for Computing Machinery, New York, NY, USA, 113–116. https://doi.org/10.1145/2493988.2494354

[28] Andrew Kurauchi, Wenxin Feng, Carlos Morimoto, and Margrit Betke. 2015. HMAGIC: Head Movement and Gaze Input Cascaded Pointing. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (Corfu, Greece) *(PETRA '15)*. Association for Computing Machinery, New York, NY, USA, Article 47, 4 pages. https://doi.org/10.1145/2769493.2769550

[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* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3173574.3173655

[30] Michael Land and Benjamin Tatler. 2009. *Looking and acting: vision and eye movements in natural behaviour*. Oxford University Press.

[31] Linnéa Larsson, Marcus Nyström, Richard Andersson, and Martin Stridh. 2015. Detection of fixations and smooth pursuit movements in high-speed eye-tracking data. *Biomedical Signal Processing and Control* 18 (2015), 145–152. https://doi.org/10.1016/j.bspc.2014.12.008

[32] Scott M Lundberg and Su-In Lee. 2017. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.), Vol. 30. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[33] Sebastián Maldonado and Richard Weber. 2009. A wrapper method for feature selection using Support Vector Machines. *Information Sciences* 179, 13 (2009), 2208–2217. https://doi.org/10.1016/j.ins.2009.02.014 Special Section on High Order Fuzzy Sets.

[34] Mark Mcgill, Aidan Kehoe, Euan Freeman, and Stephen Brewster. 2020. Expanding the Bounds of Seated Virtual Workspaces. *ACM Trans. Comput.-Hum. Interact.* 27, 3, Article 13 (may 2020), 40 pages. https://doi.org/10.1145/3380959

[35] 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* (Sydney, Australia) *(IUI '06)*. Association for Computing Machinery, New York, NY, USA, 32–38. https://doi.org/10.1145/1111449.1111464

[36] Fionn Murtagh and Pedro Contreras. 2012. Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2, 1 (2012), 86–97. https://doi.org/10.1002/widm.53

[37] Pontus Olsson. 2007. Real-time and Offline Filters for Eye Tracking. (2007), 42.

[38] Thammathip Piumsomboon, Gun Lee, Robert W. Lindeman, and Mark Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. 36–39. https://doi.org/10.1109/3DUI.2017.7893315

[39] Robert G. Radwin, Gregg C. Vanderheiden, and Mei-Li Lin. 1990. A Method for Evaluating Head-Controlled Computer Input Devices Using Fitts' Law. *Human Factors* 32, 4 (1990), 423–438. https://doi.org/10.1177/001872089003200405

[40] Najmeh Sadoughi, Yang Liu, and Carlos Busso. 2017. Meaningful Head Movements Driven by Emotional Synthetic Speech. *Speech Communication* 95 (07 2017). https://doi.org/10.1016/j.specom.2017.07.004

[41] Dario D. Salvucci and Joseph H. Goldberg. 2000. Identifying Fixations and Saccades in Eye-Tracking Protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications* (Palm Beach Gardens, Florida, USA) *(ETRA '00)*. Association for Computing Machinery, New York, NY, USA, 71–78. https://doi.org/10.1145/355017.355028

[42] Noelia Sánchez-Maroño, Amparo Alonso-Betanzos, and María Tombilla-Sanromán. 2007. Filter Methods for Feature Selection – A Comparative Study. In *Intelligent Data Engineering and Automated Learning - IDEAL 2007*, Hujun Yin, Peter Tino, Emilio Corchado, Will Byrne, and Xin Yao (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 178–187.

[43] Richard Schweitzer and Martin Rolfs. 2020. An adaptive algorithm for fast and reliable online saccade detection. *Behavior Research Methods* 52, 3 (01 Jun 2020), 1122–1139. https://doi.org/10.3758/s13428-019-01304-3

[44] Ludwig Sidenmark and Hans Gellersen. 2019. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.* 27, 1, Article 4 (Dec 2019), 40 pages. https://doi.org/10.1145/3361218

[45] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 1161–1174. https://doi.org/10.1145/3332165.3347921

[46] Ludwig Sidenmark, Diako Mardanbegi, Argenis Ramirez Gomez, Christopher Clarke, and Hans Gellersen. 2020. BimodalGaze: Seamlessly Refined Pointing with Gaze and Filtered Gestural Head Movement. In *ACM Symposium on Eye Tracking Research and Applications* (Stuttgart, Germany) *(ETRA '20 Full Papers)*. Association for Computing Machinery, New York, NY, USA, Article 8, 9 pages. https://doi.org/10.1145/3379155.3391312

[47] Ludwig Sidenmark, Dominic Potts, Bill Bapisch, and Hans Gellersen. 2021. Radi-Eye: Hands-Free Radial Interfaces for 3D Interaction Using Gaze-Activated Head-Crossing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 740, 11 pages. https://doi.org/10.1145/3411764.3445697

[48] Robert M. Steinman. 1965. Effect of Target Size, Luminance, and Color on Monocular Fixation*. *J. Opt. Soc. Am.* 55, 9 (Sep 1965), 1158–1164. https://doi.org/10.1364/JOSA.55.001158

[49] Rainer Stiefelhagen and Jie Zhu. 2002. Head Orientation and Gaze Direction in Meetings. In *CHI '02 Extended Abstracts on Human Factors in Computing Systems* (Minneapolis, Minnesota, USA) *(CHI EA '02)*. Association for Computing Machinery, New York, NY, USA, 858–859. https://doi.org/10.1145/506443.506634

[50] Robert J. Teather and Wolfgang Stuerzlinger. 2008. Exaggerated Head Motions for Game Viewpoint Control. In *Proceedings of the 2008 Conference on Future Play: Research, Play, Share* (Toronto, Ontario, Canada) *(Future Play '08)*. Association for Computing Machinery, New York, NY, USA, 240–243. https://doi.org/10.1145/1496984.1497034

[51] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. 2012. Detection of Smooth Pursuits Using Eye Movement Shape Features. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (Santa Barbara, California) *(ETRA '12)*. Association for Computing Machinery, New York, NY, USA, 177–180. https://doi.org/10.1145/2168556.2168586

[52] Daniel Vogel and Ravin Balakrishnan. 2004. Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology* (Santa Fe, NM, USA) *(UIST '04)*. Association for Computing Machinery, New York, NY, USA, 137–146. https://doi.org/10.1145/1029632.1029656

[53] Oleg Špakov and Päivi Majaranta. 2012. Enhanced Gaze Interaction Using Simple Head Gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (Pittsburgh, Pennsylvania) *(UbiComp '12)*. Association for Computing Machinery, New York, NY, USA, 705–710. https://doi.org/10.1145/2370216.2370369

[54] 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 (dec 2018), 23 pages. https://doi.org/10.1145/3287076

[55] Shanhe Yi, Zhengrui Qin, Ed Novak, Yafeng Yin, and Qun Li. 2016. GlassGesture: Exploring head gesture interface of smart glasses. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9. https://doi.org/10.1109/INFOCOM.2016.7524542

[56] Raimondas Zemblys, Diederick C. Niehorster, Oleg Komogortsev, and Kenneth Holmqvist. 2018. Using machine learning to detect events in eye-tracking data. *Behavior Research Methods* 50, 1 (01 Feb 2018), 160–181. https://doi.org/10.3758/s13428-017-0860-3

[57] Shumin Zhai, Carlos Morimoto, and Steven Ihde. 1999. Manual and Gaze Input Cascaded (MAGIC) Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Pittsburgh, Pennsylvania, USA) *(CHI '99)*. Association for Computing Machinery, New York, NY, USA, 246–253. https://doi.org/10.1145/302979.303053

[58] Yisu Zhao, Xin Wang, Miriam Goubran, Thomas Whalen, and Emil Petriu. 2012. Human emotion and cognition recognition from body language of the head using soft computing techniques. *Journal of Ambient Intelligence and Humanized Computing* 4 (02 2012). https://doi.org/10.1007/s12652-012-0107-1