A Predictive Performance Model for Immersive Interactions in Mixed Reality

Florent Cabric* Emmanuel Dubois† Marcos Serrano‡

University Of Toulouse - IRIT

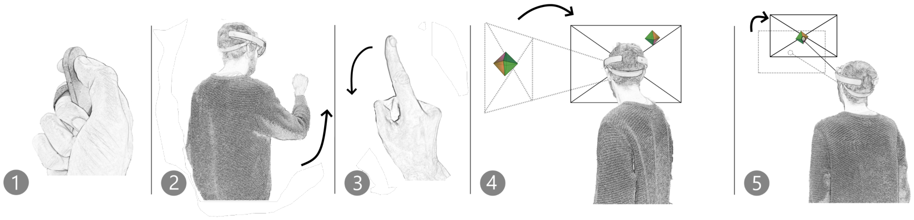

Figure 1: The five operators of our extension of the Keystroke-Level Model :1) Button Click, 2) Raise Hand, 3) Air Tap, 4) Coarse Head Point and 5) Precise Head Point.

ABSTRACT

The design of immersive interaction for mixed reality based on head-mounted displays (HMDs), hereafter referred to as Mixed Reality (MR), is still a tedious task which can hinder the advent of such devices. Indeed, the effects of the interface design on task performance are difficult to anticipate during the design phase: the spatial layout of virtual objects and the interaction techniques used to select those objects can have an impact on task completion time. Besides, testing such interfaces with users in controlled experiments requires considerable time and efforts. To overcome this problem, predictive models, such as the Keystroke-Level Model (KLM), can be used to predict the time required to complete an interactive task at an early stage of the design process. However, so far these models have not been properly extended to address the specific interaction techniques of MR environments. In this paper we propose an extension of the KLM model to interaction performed in MR. First, we propose new operators and experimentally determine the unit times for each of them with a HoloLens v1. Then, we perform experiments based on realistic interaction scenarios to consolidate our model. These experiments confirm the validity of our extension of KLM to predict interaction time in mixed reality environments.

Index Terms: Human-centered computing—Human-Computer Interaction—HCI theory, concepts and models

1 INTRODUCTION

Over the past few years, the advent of immersive mixed reality based on head-mounted displays, hereafter referred to as mixed reality, has pushed the visualization and interaction research fields, bringing multiple novel advances and applications. Head-mounted displays (HMDs) now include spatial registration, to place interfaces anywhere in our environment, as well as mid-air gestures and gaze tracking to perform spatial pointing. As a result, developing immersive mixed reality systems requires to consider the effects of the spatial layout of the interface, which can be situated outside the user's Field of View (FoV). One problem for researchers and designers is to assess the time required to perform such pointing and validation tasks. Developing and testing immersive mixed reality HMD-based applications is tedious, hence the usual combination of prototype-based development and user studies can be inefficient in the context of immersive HMD-based mixed reality interaction [1].

To facilitate the exploration of the design possibilities in such context, we present an extension of the well-known Keystroke-Level Model (KLM) predictive model [6]. KLM allows designers to predict interaction times from atomic interactions called operators, and identify usability issues during the design phase rather than during the development phase [4] Although KLM has been subject to numerous extensions for specific domains (such as smartphone-based interaction), little attention has been paid to modeling interactions in the context of mixed reality environments. And yet, such context involves specific behaviors such as moving the head to point at the correct piece of information in the immersive space, or bringing the hand in the HMD's field of view to perform a mid-air validation that can be detected and recognized by the HMD. The originality of our work is to specifically focus on the design of an extension of KLM, which will be used to model interaction to perform mixed reality tasks such as pointing and selecting digital objects.

To this end, we first propose a set of operators, including new mixed reality operators and operators from the literature, that are keys to model the required atomic interactions relevant in the context of mixed reality environments. Regarding our mixed reality operators, we experimentally measure their respective unit times, i.e. the time required to perform the corresponding action. In three final user studies, we measure the time required to perform a task that combines the mixed reality KLM operators, and compare this measured completion time to the predicted time computed with our extension of the KLM model. These final studies validate our contributions since the differences between the observed and predicted times are on average less than 5%, which is compliant with a KLM approach [33].

Our contributions are 1) the identification of relevant KLM operators to describe interactions with mixed reality, 2) the definition of appropriate unit times for these operators through 2 sessions of user studies and 3) the experimental consolidation of our model.

2 RELATED WORK

HCI evaluation traditionally relies on user studies, which are sometimes challenged by the complexity of tasks, the multiplicity of interaction modalities or particular usage contexts. Alternatively,

---

*e-mail: florent.cabric@irit.fr
†e-mail: emmanuel.dubois@irit.fr
‡e-mail:marcos.serrano@irit.fr

model-based evaluation [22] allows designers to evaluate a system or an interaction technique prior to its implementation [19–21]. In this section, we review existing works on the Keystroke-Level Model (KLM) and its extensions. KLM is one of the best performing models when it comes to predicting completion time of a user interaction [22].

2.1 KLM

The Keystroke-Level Model (KLM) [6] is one of the GOMS family models and is the easiest model of the GOMS-family (Goals, Operators, Methods, Selection) [7] to use for application designers unlike other models (e.g. cognitive walkthrough [34], cognitive architecture models [5] or task network models [9, 24]), such approaches model a routine task (i.e. a task performed by the user without error) to obtain a prediction of the completion time for the task.

In comparison to the initial GOMS model, KLM focuses on the execution of the task once it has been planned, i.e. independently from cognitive considerations and user's strategies. This contributes to make KLM the easiest model of the GOMS-family [7] to use for application designers. . With KLM, modeling a task requires decomposing the task into a sequence of atomic actions such as key press, pointing, mental acts, homing, etc. Each action is represented by an operator which is characterized by a unit time. The unit time corresponds to the duration of the corresponding atomic action performed by the user. Consequently, the time to realize the task is equal to the sum of the unit time of each operator modeling the task. KLM model was initially limited to desktop environments, but the emergence of new devices or user's needs has led researchers to design extensions to the KLM model.

2.2 KLM Extensions

One of the reasons to extend the KLM model is to evaluate new technologies or applications that were not considered in the original KLM model [1]. Extending a KLM can take different forms: new operators, new heuristics or an adaption of existing operators. The advent of smartphones led researchers to propose multiple extensions for mobile phone-based interaction [11, 25, 31, 35]. Among them some works explored extensions of KLM by adding new operators (e.g. Tilt [35]) or Zoom [11]). Others proposed adapting operators from the original KLM for mobile phone-based interaction (e.g. Keystroke or Pointing [17]) and resulted into an extended model able to predict a completion time for a task with a margin of error between -15% and +8%. KLM extensions also addressed the use of mobile devices in different contexts of use requiring specific interactions and thus specific operators : for example gaming applications [25] requires numerous repetitive touch on a located area.

In-vehicle systems have also been widely investigated [14, 27, 32, 36]. For instance, Schneegaß et al. [36] model a new formula to integrate the separation between the main driving task and the secondary interaction tasks. Green [14] added new operators to evaluate navigation tasks according to the "15-Second" rule (i.e. the navigation task should not take more than 15 seconds). Moreover, several works explored extensions of KLM for new input devices on in-vehicle systems (e.g. control type and menu structure [27] or remote control [32]).

More relevant for mixed reality environments, other extensions explored Natural User Interactions (NUIs), i.e. interaction with hand or arm gestures only. For example, the work of Erazo et al. [13] aims to describe interaction with hand gesture interfaces. The gestures are detected with a Kinect. They use Gesture unit (G-unit [28]) and adapt a KLM model to estimate the time taken for each part of the gesture. For instance, they define a Pointing (P) operator to model the time needed to move the hand from a starting point to an ending point. Also, they designed new heuristics to place their Mental operators in their task modeling. For interested readers, Al-Megren et al. [1] propose a recent state of the art on KLM extensions. However, to our knowledge and according to Al-Megren state of the art, there is still a lack of KLM extensions for interactions for mixed reality applications.

3 OUR KLM EXTENSION FOR MIXED REALITY

In this section, we present our KLM extension for interaction in mixed reality. In particular, we aim to model Button Click, Raise Hand, Air Tap, Coarse Head Point and Precise Head Point in mixed reality environments (cf. Fig. 1). These tasks are fundamental components of any mixed reality interaction; their definition and characterization are therefore essential. We first present the mixed reality operators before introducing an inherited operator that we reuse from previous literature.

3.1 New operators for mixed reality interaction

To investigate the validation in mixed reality environments, we considered two different techniques, commonly used in such environments: pressing on a physical button and performing a mid-air validation gesture. We therefore defined three operators: Button Click, which corresponds to the atomic action "pressing a physical button", and, Raise Hand and Air Tap, that corresponds to two atomic actions often performed sequentially to allow the detection of a mid-air validation gesture in the FoV of the HMD. As already mentioned, moving the head to point at the correct piece of information in an immersive space is also one of the recurrent actions required in such environment. We therefore also investigated two forms of pointing in the mixed reality environment: Coarse Head Point and Precise Head Point. We further detailed these five operators in the following subsections. These operators are illustrated in Fig. 1.

3.1.1 Button Click (**Bc**)

Performing repetitive mid-air gestures could be cumbersome and tiring. Several mixed-reality HMDs (e.g. HoloLens or Magic Leap One) are thus delivered with a dedicated handheld device to perform validation actions. Our operator Button Click (**Bc**) models the gesture to perform a click with such a dedicated device.

3.1.2 Raise Hand (**Rh**)

Keeping the hand in the air can be painful and prone to fatigue ("gorilla-arm effect" [16]). Therefore, one recurrent action in Mixed Reality contexts is to raise the hand so that the HMD can detect it with its embedded technologies and recognize the gestures. The Raise Hand (**Rh**) operator models the gesture needed for the user to bring his hand from a rest position to a position detected by the headset.

3.1.3 Air Tap (**At**)

Selecting a virtual object or clicking on a virtual button are two major actions useful in mixed reality environments to select an object or to activate a command. Many HMDs offer mid-air gestural validation to select virtual elements. Our operator Air Tap (**At**) models such mid-air gestural validation. In this work, we decide to model the built-in HoloLens gesture gesture (present in the HoloLens v1 and v2), which consists in lowering the index finger and raising it again, due to its robustness and simplicity to perform.

3.1.4 Coarse Head Point (**Chp**)

Mixed reality devices have a restricted FoV in comparison to the human FoV. Typically, the HMD offer a FoV of 30° for Hololens or 90° for Meta2, while the human binocular FoV is about 120°. Therefore, before being able to read or select an element, this element must be brought into the FoV of the HMD. This is what is depicted by this operator Coarse Head Point and corresponds to searching for a target outside the HMD's FoV, another very common task in mixed reality applications. The Coarse Head Point (**Chp**) operator thus describes the gesture of a user to look for an object outside the

HMD's FoV. This action involves large and multidirectional head movements.

3.1.5 Precise Head Point (**Php**)

Once an element is visible in the FoV, a precise selection of this element, potentially among several elements, has to be done before being able to interact with it. The Precise Head Point (**Php**) operator models the task that consists for a user in pointing at an object already visible in the HMD's FoV with a cursor. This operator involves slight head movements to bring the cursor from a starting point to an ending point (i.e. the target). Precise Head Point in most HMDs is viewport-based [8]: a cursor is placed in the center of the FoV and remains in the center when users move their head.

3.2 Adaptation of an existing operator: Simple Reaction (Ms) and its heuristics

Initially, Card et al. [6], modeled a time required between 2 different physical atomic actions: the **M** operator. This operator can model different cognitive tasks (i.e. visual search, plan for future actions, etc.). MacKenzie [26] refined this operator by splitting it into 5 different mental operators: simple reaction (**Ms**), physical matching (**Mp**), name matching (**Mn**), class matching (**Ml**) and visual search (**Mv**). The unit time for each operator was experimentally measured. We chose to rely on them in our extension of KLM, especially the simple reaction (**Ms**) operator and its unit time (277 ms) which is the only one relevant for pointing and validating tasks.

The insertion of the different mental operators in a KLM prediction is based on heuristics. The KLM from Card et al. [6] proposed five heuristics to place the original Mental Act (**M**). First, add an **M** operator in front of each physical operator (Rule 0). Then, remove **M** operators in front of any operators that can be fully anticipated (Rule 1). In a cognitive unit delete all **M** but the first (Rule 2). The next two rules concern command-line interfaces (Rule 3 and 4) and are not relevant for our extension of KLM.

Kieras [21] has proposed some guidelines to improve the use of these original heuristics. Among these guidelines, the designers should focus on the number of mental operators rather than their placement. This specific guideline is taken into account in existing works on KLM extensions [13, 17] while others are domain specific and irrelevant for our context. Based on these guidelines and the original heuristics, four rules will guide the use of Mental operators in our KLM extension:

- Rule 0: Place **Ms**'s in front of any of our operators (Air Tap's, Button Click's, Raise Hand's, Precise Head Point's and Coarse Head Point's).
- Rule 1: If an operator following an **Ms** is anticipated in the operator before **Ms**, delete the **Ms**. For instance, when the user must point and select a target, remove the **Ms** between the pointing and the selection operators.
- Rule 2: If a string of **Ms**+**Bc**'s or **Ms** + **At**'s belongs to a cognitive unit then delete all **Ms**'s but the first. For instance, if the user needs to perform a multiple Button Click **Bc** (or Air Tap **At**), remove all **Ms** between each **Bc** (or **At**).
- Rule 3: If unsure, emphasize more the number than the placement of the occurrences of the **Ms** operator

These five operators are introduced as an extension of the KLM model since they correspond to atomic actions. In the next section, we define, through user studies, the unit time associated for each mixed reality operators.

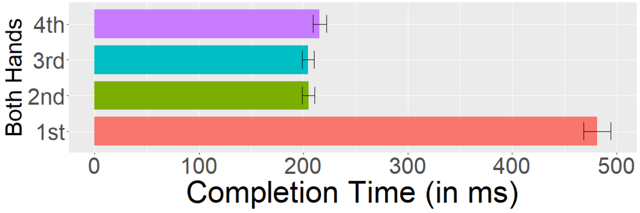

Figure 2: Button Click (**Bc**) observed time with (left) dominant and (right) non-dominant hand for each click

4 USER STUDIES: DEFINING A UNIT TIME FOR THE NEW OPERATORS

After having introduced the new operators of our KLM extension, we experimentally measure the unit time for each one of them, i.e. the predicted duration of the corresponding user's action.

4.1 Participants and apparatus

We performed two separate sessions to determine the operators' unit times: one for Button Click operator, the other for the Air Tap, Raise Hand, Coarse Head Point and Precise Head Point operators. 12 participants took part in the first session (3 females and 9 males aged 25 to 43, mean 30.2). Among them, 5 are graduate or undergraduate, 2 are Ph.D. fellows, 3 are engineers and 2 are lecturers in our computer science lab. 12 participants took part in the second session (3 females and 9 males aged 23 to 32, mean 27.4). Among them, 10 are graduate or undergraduate students and 2 are engineers. 50% of these participants performed the two sessions. This number of participants is consistent with other studies on KLM [17, 31, 36].

For these two sessions, we used a HoloLens HMD v1 and its Clicker (i.e. the selection dedicated device). The experiments took place in a controlled environment with no natural light. A video illustrating each task was shown to the participants before the experiment. Due to the fact that the experimenter cannot see what the participant sees in the HMD, the video overcomes the difficulty of explaining a task. Participants could ask any questions while the video was playing. Following this video, participants performed an official HoloLens tutorial called "Learn Gestures", in which they learned how to perform the basic gestures for HoloLens' interactions: Raise Hand, Air Tap and Precise Head Point. All these studies were carried out in such a way that the sanitary measures imposed by COVID-19 were scrupulously applied.

4.2 Data collection and analysis

All the time measures collected in these experiments are captured by the HoloLens and correspond to system events managed by the Mixed Reality Toolkit (MRTK) [30]. We computed geometric means instead of the arithmetical mean, as it better suits data with a long-tailed distribution such as completion time [12], with 95% confidence intervals. All scripts are available online[1].

4.3 Button Click operator

4.3.1 Task

To measure the time required for clicking with a dedicated device, we asked participants to perform four successive clicks. A feedback was displayed between each Button Click (**Bc**) gesture to indicate the number of clicks already performed by the participants. To limit cognitive or physical users' actions, the participants had to look at a timer displayed in front of their eyes and wait for the 3s countdown to start the trial. The same task was performed with both the dominant and non-dominant hands. Participants first went through a training session composed of five trials, before completing the experiment

---

[1] https://osf.io/pjyhk/?view_only=bb36af530d71431094c4c76eae7ed1b1

which included 15 trials. We recorded the completion times from the end of the countdown to the first click and between each click. We collected 4 Button Click x 2 hands x 15 trials = 120 Button Clicks per participant.

4.3.2 Results

In a first analysis, we found no difference between each hand. We thus decided to compute the geometric mean for dominant and non-dominant hand together. We now investigate these results for each click independently.

The first click was performed in 481ms (CI[468,494]). This first click took much more time than the others (2nd click: 205ms, CI[199,211]; 3rd click: 205ms, CI[199,210]; 4th click: 213ms, CI[209,221]). This can be explained by the fact that, as underlined by MacKenzie [26], the user takes a simple reaction time to react after the end of the countdown. This simple reaction (the **Ms** operator) is on average 277ms. If we subtract this reaction time to the first click, it remains that the first click took 204ms CI[191,218] (cf. Fig. 2) which is in line with the time required to perform the three other clicks. According to our heuristics (Rule 2 cf. Sect. 3.2), there is no need to subtract the simple reaction unit time from the others clicks.

After having adjusted the time measured for the first click, we computed the geometric mean for each click for both dominant and non-dominant hand. The mean correspond to the unit time for the Button Click (**Bc**) operator and equals 207ms CI[202,213].

4.4 Raise Hand Operator

4.4.1 Task

To measure the time required for raising the hand (**Rh**), before performing the Air Tap (**At**) gesture, we asked participants to bring their dominant hand within the FoV of the HoloLens. Before starting each trial, participants had to place their hand along their thigh. To limit cognitive or physical users' actions, the participants had to look at a timer displayed in front of their eyes and wait for the 3s countdown to start the trial. Indeed, performing actions (cognitive or physical) other than the action corresponding to the operator **Rh** could lead to a bias in the unit time measured. The participants completed 5 training trials, followed by 15 study trials. We recorded the time corresponding to **Rh** (i.e. between the end of the countdown and the hand being detected by the HoloLens). We collected 15 Raise Hand gestures per participant.

4.4.2 Results

The mean time from the end of the countdown to the moment the HoloLens detects the hand is 899ms (CI[861,938]). However, as underlined by MacKenzie [26], and already detailed in the previous section, the user takes a simple reaction (**Ms**) time to react after the end of the countdown. Therefore, we subtract the reaction time of 277ms from the mean measured time. Consequently, the unit time corresponding to the Raise Hand operator is 623ms CI [586,662].

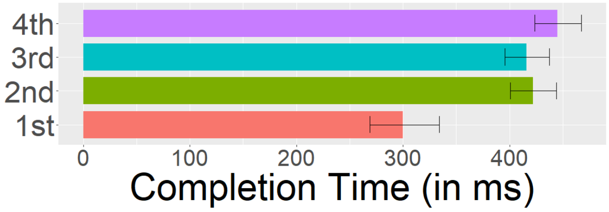

Figure 3: The four Air Taps completion times in ms with 95% CI

4.5 Air Tap operator

4.5.1 Task

Regarding the Air Tap (**At**) operator, participants had to perform four successive Air Tap gestures after only one Raise Hand gesture. A feedback was displayed between each gesture to indicate the number of Air Tap gestures already performed by the participant. The time recorded for one Air Tap gesture corresponds to the time required to lower the index finger and raise it again. The participants completed 5 training trials, followed by 15 study trials. We collected 4 Air Taps x 15 trials = 60 Air Taps per participant.

4.5.2 Results

We investigate completion time by considering each of the four consecutive Air Tap gestures independently. As illustrated in Fig. 3, we observe that the first Air Tap gesture tends to take less time than the three following ones (300ms, CI[269,334]). We believe that this difference is due to the fact that participants started the first Air Tap while raising their hand. We thus decided to exclude the first Air Tap gesture from our calculation. As a result, the mean time of the Air Tap is 427ms CI[408,448]. Unlike the **Bc** or **Rh** operators, the action of performing the three last consecutive Air Tap gestures does not require any simple reaction time (cf. Rule 2 Sect. 3.2), thus there is no need to subtract the unit time corresponding to a simple reaction (**Ms**). Therefore the unit time characterizing our **At** operator is 427ms CI[408,448].

4.6 Coarse Head Point operator

4.6.1 Task

Bringing the FoV around a target to visualize it in a mixed reality environment consists in rotating the head until the target enters the HMD's FoV. The time required to bring the FoV on the target depends on the angular distance from the target to the initial position of the center of the FoV. According to [37], the maximum comfortable head rotation is 60° (without shoulder rotation and body inclination). Thus, we defined three angular distances to the target outside the FoV of the HoloLens (45°, 60° and 75°). Moreover, as underlined in [37] it is important to distinguish vertical from horizontal head movements. In our study, we therefore positioned targets in 8 directions corresponding to the possible combinations of horizontal and vertical head movements: two vertical (North, South), two horizontal (East, West), and four diagonal movements (North East, North West, South East and South West, cf. Fig. 4). Finally, and according to HoloLens guidelines [29], 7.5cm targets were placed on a 2m radius sphere centered on the user's eyes (i.e. an angular size of 2.14 degrees). Each target is placed in one of eight directions. Combining the 3 angular distances with the 8 directions gives 24 possible target positions around the FoV.

To start a trial, participants had to wait for a 3s countdown placed on a starting object displayed in front of their eyes, when looking straight ahead. Then, an arrow pointed out the direction to follow to find the target, and hence bring the FoV on the target. More precisely, participants had to put the target inside a circle displayed at the center of the FoV and inscribed into the FoV, i.e. so that its diameter fits exactly the height of the FoV. This was done to remove any effect due to the difference between the FoV's width and height. The radius of this circle (inscribed in the FoV) has an angular distance of 8.75° which is the half of the height of the FoV of the HoloLens (17.5°) and within the range that can be perceived by the human eye (i.e. between 5° and 10° [10,23]). The trial automatically ended once the target had been in this circle for 500ms. Targets are displayed one after another.

This part of the study followed a 3x8 within-subject design with angular distance (45°, 60°, 75°) and direction (2 horizontal, 2 vertical, 4 diagonal) as factors. Each block corresponded to one angular distance and included eight targets (one in each direction) randomly

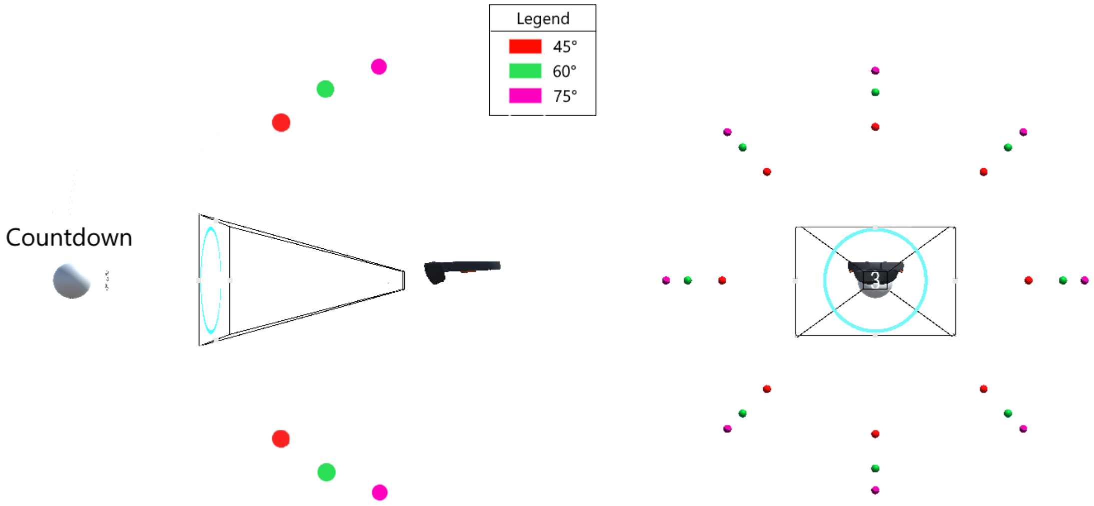

Figure 4: Setup Coarse Head Point with (red) 45 ° targets, (green) 60 ° targets,(pink) 75 ° targets, (white) the countdown object and (blue) the circle inscribed in the FoV. On the left, the north and south target in a side view and on the right the whole setup with the 24 targets in a front view.

presented to the participant. Blocks were counterbalanced over participants. Each block was repeated three times, the first one serving as training. In total, we collected 3 angular distances x 8 directions x 2 repetitions = 48 trials per participant.

The completion time was measured between the moment the HoloLens gaze cursor left the object on which the countdown was displayed and, the moment the target entered the circle of view (i.e. we removed the 500ms of trial validation).

4.6.2 Results

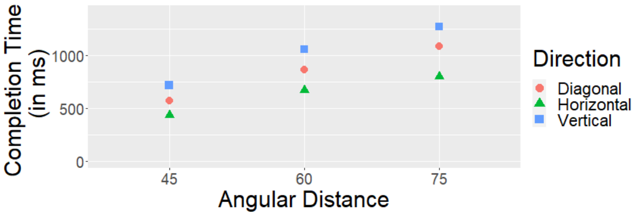

Figure 5: Time completion for Coarse Head Point according to direction and angular distance

First, we observed that the angular distance has a clear impact on the time required to bring the circle inscribed in the FoV around the target: time clearly increases with the angular distance (45°: 561ms CI[518,607]; 60°: 852ms CI[796,913]; 75°: 1044ms CI[969,1125]). The direction also impacts this completion time (986ms CI[907,1072] for vertical movements, 613ms CI[560,671] for horizontal movements and 810ms CI[757,866] for diagonal movements). Fig. 5 illustrates these results and shows that these results remain true when considering these two factors simultaneously.

As ([36], [25]), we extracted from these results a set of 9 geometric means (cf. Table 1) and that can be used as unit time for the Coarse Head Point operator. These values depend on the direction (d) and the angular distance (a) and will be useful to help the designer in deciding where to display widget, window, tools or 3D elements in a working space. The potential simple reaction time of the user is not included in these values as the measure of the completion time started when the gaze cursor left the object on which the countdown was displayed, after the user had already reacted.

Table 1: Means, in ms, for Coarse Head Point according to direction (d) and angular distance (a) : (**Chp(d,a)**)

	Horizontal	Diagonal	Vertical
45	431 CI [365,510]	567 CI[510,629]	715 CI[617,829]
60	668 CI [595,752]	865 CI[783,957]	1054 CI[936,1187]
75	797 CI [695,914]	1083 CI[970,1209]	1270 CI[1121,1441]

4.7 Precise Head Point operator

4.7.1 Task

This study aims at measuring the time required to point at a target already present in the FoV with the gaze, an action corresponding to our last operator: Precise Head Point (**Php**). We positioned eight targets on the circle inscribed in the FoV as defined for the previous part of the study, according to the eight cardinal directions (Fig. 6) and within the range that can be perceived by the human eye [10, 23]). To start a trial, the participants had to wait for a 3s countdown displayed in front of their eyes as in the previous part of the study. Then, they had to bring the gaze cursor of the HoloLens on the target and stay 500ms to validate the trial. The completion time was measured between the moment the HoloLens gaze cursor left the area where the countdown was displayed and, the moment it entered the target, thus avoiding the 500ms of dwell validation.

4.7.2 Results

As opposed to the action corresponding to the Coarse Head Point (**Chp**) operator, results concerning the Precise Head Point (**Php**) operator do not allow to identify an impact of the direction on the

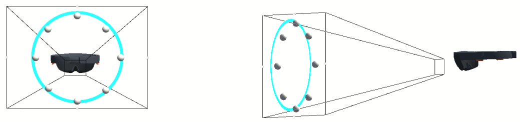

Figure 6: Setup for Precise Head Point task with the FoV and the targets (in white)

completion times required to point at a target already present in the FoV (Horizontal = 452ms, CI[369,554]; Diagonal = 439ms, CI[378,506]; Vertical = 358ms, CI[299,429]) cf. Fig. 7). For the same reason than with the Coarse Head Point operator, there is no need to subtract a "simple reaction" time. From these results the unit time for Precise Head Point (**Php**) is 419 ms.

In addition, we collected the number of times the cursor entered and exited the target. Of the 192 trials, 105 were carried out without exiting the target (54%), 76 were carried out by exiting the target once (39%) and only 11 (5.7%) required more than two entries on the target before final selection.

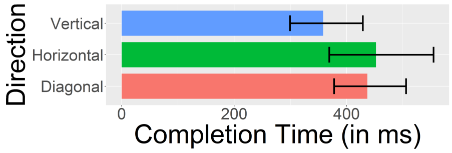

Figure 7: Precise Head Point completion time according to direction (in ms) with 95% CI

4.8 Summary

Through these different experiments, we defined unit time values for the mixed reality operators of our KLM extension for mixed reality interaction: Button Click (207 ms), Raise Hand (623 ms), Air Tap (427 ms), Coarse Head Point (9 values according to angular distance and direction), and Precise Head Point (419 ms). All the operators and their unit time associated are summed up in Table 2.

These operators and their unit times associated now need to be consolidated within ecological tasks. To this end, three user studies were designed to confront our model with different combinations of operators.

Table 2: Our extended KLM

Name	Time Unit (in ms)	From
Simple Reaction (**Ms**)	277	[26]
ButtonClick (**Bc**)	207 CI [202, 213]	Mixed Reality
Raise Hand (**Rh**)	623 CI [586,662]	Mixed Reality
Air Tap (**At**)	427 CI [408,448]	Mixed Reality
Coarse Head Point (**Chp**)	From 431 to 1270 cf. Table 1	Mixed Reality
Precise Head Point (**Php**)	419 CI [369,554]	Mixed Reality

5 CONSOLIDATION STUDIES

The three consolidation studies correspond to different combinations of operators required to performed different tasks that are common in mixed reality applications: 1) a pointing task where targets are placed outside the initial FoV; 2) a multiple selections task where targets are placed outside the initial FoV; 3) a multiple selections task where targets are placed inside the initial FoV. The operators involved in each task are summarized in Table 3. Each of these three consolidation studies consists in comparing an experimentally measured completion time of the task with a predicted time of the task, based on the use of the unit times associated to existing and newly inserted KLM operators.

5.1 Participants and apparatus

These three consolidation studies took place during two sessions. The first session only concerned the pointing task (study 1) and the second session concerned the two multiple selections (in FoV and out FoV) tasks (studies 2 and 3). Twelve participants (3 females and 9 males aged 25 to 43, mean 30.2) took part in the first consolidation study. Among them, 5 are graduate or undergraduate, 2 are Ph.D. fellows, 3 are engineers and 2 are lecturers in our lab. All participants of this first study were participants of the first session of operators studies. Twelve participants (3 females and 9 males aged 26 to 51, mean 31.2) performed the consolidation studies 2 and 3. Among them, 8 are graduate students, 2 are Ph.D. fellows and 2 are engineers. 3 of them did not participate in any operators' study, 3 participated in the first session of operators studies, 4 participated in the second session and 2 participated in both sessions. This number of participants is consistent with other studies on KLM [17,31,36].

Table 3: Consolidation studies and the operators involves for each study

Operators	Pointing out of the FoV	Multiple selections out of the FoV	Multiple Selections in the FoV
ButtonClick (**Bc**)		X	
Raise Hand (**Rh**)			X
Air Tap (**At**)			X
Coarse Head Point (**Chp**)	X	X	
Precise Head Point (**Php**)	X	X	X
Simple Reaction (**Ms**)	X	X	X

For these three studies, we used the same apparatus as in the previous experiments for the unit times.

5.2 Data collected and data analysis

All the time measurements collected in these experiments are captured by the HoloLens and correspond to system events managed by the Mixed Reality Toolkit (MRTK) [30]. We computed geometric means for each trial for each participant [12]. As already adopted in previous extensions of KLM, we considered our model validated if difference between the predicted times and the observed times are in the 20% error criterion from Olson and Olson [33].

5.3 Consolidation Study 1: Pointing Out of the FoV

A very common task in mixed reality environment is to point at a target that is initially outside the FoV. For example, in an augmented maintenance task or mixed reality surgery, the user (e.g. the technician or the surgeon) can point and select commands in a mixed reality app displayed out of his field of work. However, the user has tools in his/her hands thus this interaction must be hands-free. With a HMD, in mixed reality pointing can be performed with head movement. We decided to implement the validation with a dwell time of 500ms (e.g. as [2]). This task is thus composed of a combination of a Coarse Head Pointing task, our **Chp** operator, to bring the FoV around the target, and a Precise Head Pointing task, our **Php** operator, to select the target that is now in the FoV.

5.3.1 Study design and task

The task took place in 2 steps: 1) searching for the target in the direction indicated by an arrow and placing it inside the circle inscribed in the FoV (**Chp** operator) and 2) pointing with the HoloLens cursor at the target (**Php** operator). We used the same protocol as for the Coarse Head Point task with 24 targets (eight targets placed at three angular distances). In total, we collected 3 angular distances x 8 directions x 2 repetitions = 48 trials per participant.

5.3.2 Task Modelling

Executing the task involves two operators, **Php** and **Chp**. In addition, KLM model proposes heuristics to place simple reaction operators between two physical actions. In our task modeling and according to our heuristics (Rule 0), a mental operator must be added before the **Php** and the **Chp** operators. However, the completion time is only calculated from the moment when the user starts to make the gesture represented by the Coarse Head Point operator. The first reaction time (Ms) of the user is therefore not contained in the execution time and therefore in our model

In the previous study, we found that **Chp** depends on the angular distance of the target and the direction of the head movement. We defined nine different units times for **Chp(d,a)**, where d is the direction (Horizontal, Vertical or Diagonal) and a is the angular distance (45°, 60° or 75°). Thus, nine predicted times (pt), one for each combination of direction and angular distance, were computed with our KLM extension (cf. Sect. 4.6.2): pt(d,a) = **Chp(d,a)** + **Ms** + **Php**. For instance, for a target placed on the diagonal and at an angular distance of 60°, the predicted time is pt(diagonal, 60°) = 865 + 277 + 419 = 1561 ms.

5.3.3 Comparison between experimentally measured and predicted time

In the Table 4, we report the mean completion time of the task with 95% CI for each of the 3 directions and angular distances. In each cell, we compute the difference (in ms and percentage) between the predicted time and the means of observed times.

For the nine predictions, the percentage of difference ranges from 0% (6ms difference for horizontal 60) to 10% (135ms difference for diagonal 60) (cf. Table 4). On average, we found an error of 4%. All these results consolidate our model for the sequence of the nine **Chp**, the **Ms** and the **Php** unit times.

5.4 Consolidation Study 2: Multiple selections out of the FoV

Selecting different targets is another very common task in mixed reality applications. For instance, a complex 3D model can be displayed inside the FoV and different commands useful to modify, update or manipulate the model are available on a separate panel outside the initial FoV. In such context, the user must select the 3D model search the commands by looking away from its initial FoV, and finally select one of them before performing these steps again. In comparison to just "pointing out the FoV" (cf. Sect. 5.3) this study includes the final validation at each step performed with a Button Click and increases its external validity as it considers a sequence of actions (pointing and selecting a target outside the FoV).

5.4.1 Study design and task

The task starts with the selection of a first target displayed in the center of the FoV (cf. Fig. 8 left), then searches for a target placed at 45° in a horizontal direction (**Chp**(horizontal,45)), selects it and comes back to the first to select it (cf. Fig. 8 right), and so on. The selection is made through a sequential use of a Coarse Head Point **Chp**(horizontal,45), a simple reaction (**Ms**), a Precise Head Point (**Php**) and a validation step based on a Button Click (**Bc**).

During the task, an arrow placed above the targets indicated the direction of the next target. As in the previous studies, the participants had to look at a 3s countdown displayed on the first target in front of their eyes to start the trial. The targets measured 7.5 cm (angular size = 2.14 degrees) and were placed in a 2m sphere. Targets that the user had to select outside the initial FoV were randomly positioned to the left or right of the initial FoV. The trial ended when the last target was selected. The participants completed 6 trials as a training phase, followed by 18 trials composing the study. We collected the completion time of 18 trials per participant.

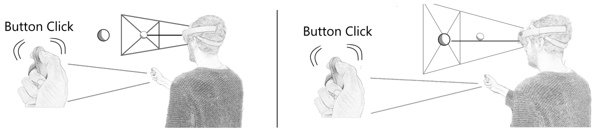

Figure 8: Setup of one trial for selection outside the FoV task with (left) the first target placed in the center of the initial FoV and the user performs a Button Click and (right) the user's head rotated 45° to the left to point at the second target and s/he performs a Button Click.

As we already established the validity of the unit time defined for the sequence of **Php** and **Chp** operators in any direction and angular distance, we decided to consider only one direction and angular distance for the Coarse Head Point operator in this study.

5.4.2 Task modelling

This five targets selection task requires numerous repetitions of Coarse Head Point (**Chp**(horizontal,45)), Precise Head Point (**Php**) and Button Click (**Bc**). The first selection only includes a Button Click (**Bc**) and the following four include (in this order): 1) **Chp** 2) **Php** and 3) **Bc**. According to rule 0 from heuristics in Sect. 3.2, a simple reaction (**Ms**) operator is added before each operator.

Then, the **Ms** between a **Php** and a **Bc** and between the **Bc** and **Chp** are deleted because both **Bc** and **Chp** can be fully anticipated during their previous action (rule 0 Sect. 3.2). Then, the following four selections are modelled with a **Chp**(horizontal, 45), a **Ms**, a **Php**, and a **Bc**. Finally, the predicted time is: 1 x (**Ms** + **Bc**) + 4 x (**Chp**(horizontal, 45) + **Ms** + **Php** + **Bc**) = 1 x (277 + 207) + 4 x (431 + 277 + 419 + 207) = 5 820ms.

5.4.3 Comparison between observed and predicted times

As detailed above, the predicted time for the entire trial is 5 820ms. The geometric mean of all observed completion times is 5 783ms (CI[5 641,5 928]). Thus, the difference between the observed times and the predicted time is 37ms (i.e. less than 1% error between the predicted and the observed times). This result strongly validates different elements of our model in the context of a task involving a multiple selection of targets placed outside the FoV.

5.5 Consolidation Study 3: Multiple selections in the FoV

Most of mixed reality applications display several objects at the same time. Thus, selecting different objects inside the FoV is a third rather common task in such applications. For instance, in the Learn Gestures tutorial embedded in the HoloLens, holograms must be selected to learn the Air Tap gesture. In this third consolidation study, we aim to validate our model when the interaction involves a multiple selection task of targets displayed exclusively inside the FoV. Another difference with the Consolidation Study 2 is that we explored the use of the Raise Hand and the Air Tap instead of the Button Click to confront them in ecological interactive situations.

5.5.1 Study design and task

The task is a selection of five successive targets placed at an angular distance of 8.75° from each other and from left to right. To avoid any learning effect, a trial was composed of two targets placed at

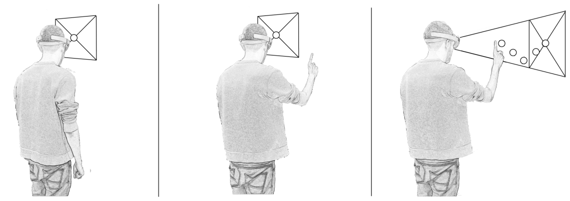

Figure 9: Setup of one trial for selection inside the FoV task with (left) the participant pointed at the first target with the HoloLens, (center) the participant raising his hand and his finger to perform the Air Tap validation and (right) the selection of the fifth (the last) target.

Table 4: Completion time (CT) with 95% CI (in ms), the Predicted Time (PT) computed with our model and the difference (in ms and in percentage) between the mean completion and the predicted time.

	Horizontal			**Diagonal**			**Vertical**		
	CT	PT	Diff	CT	PT	Diff	CT	PT	Diff
45	1155 [1040,1283]	1127	+28 (2.5%)	1299 [1218,1385]	1263	+36 (2.8%)	1491 [1295,1716]	1411	+80 (5.4%)
60	1229 [1109,1360]	1364	-135 (-10%)	1567 [1462,1680]	1561	+6 (0%)	1601 [1449,1768]	1750	-149 (-9.3%)
75	1440 [1303,1592]	1493	-53 (-3.7%)	1796 [1665,1938]	1779	+17 (1%)	1924 [1761,2102]	1966	-42 (2.1%)

the northeast of the previous target and two targets placed at the southeast (cf. Fig. 9). To select a target, the participants must bring the gaze cursor inside the target (**Php**) and perform an Air Tap (**At**). At the beginning of the trial, only the first target was visible. The next target appears on the right of the current target as soon as the current target is selected. Once the last target was selected, the countdown reappeared to start a new trial. During the countdown, we asked participants to lower their arm in order to avoid fatigue. The 7.5 cm targets (angular size: 2.14 degrees) were displayed at 2 meters away from the user.

The participants completed 6 trials for the training and repeated three times the 6 trials for the study. Thus, 6 x 3 = 18 observed completion times were collected per participant. We collected the observed times from the first pointing (i.e. when the gaze cursor enters on the target) to the selection of the last target.

5.5.2 Task Modelling

We model the first selection with a Raise Hand (**Rh**) and an Air Tap (**At**) because we record the time when the cursor enters the first target. The next four selections were modeled with a Precise Head Point (**Php**) and an Air Tap validation (**At**). There is no Raise Hand because the hand remains in the FoV between two targets.

After the addition of all **Ms** operators and according to our heuristics (Rule 1), we delete the **Ms** between the **Rh** and the **At** operators (first selection) and between **Php** and **At** operators (last four selections) because the action of Air Tap (**At**) can be fully anticipated during the Raise Hand (**Rh**) and the Precise Head Point (**Php**) gestures. Thus, the predicted time is: 1 x (**Rh** + **At**) + 4 x (**Ms**+ **Php** + **At**) = 1 x (623 + 427) + 4 x (277 + 419 + 427) = 5 542ms.

5.5.3 Comparison between observed and predicted times

As detailed above, the predicted time for the entire trial is 5 542ms. The geometric mean of all observed times is 5 271 ms (CI[5101,5446]). Thus, the difference between the mean observed and the predicted time is 271 ms (i.e. 4.8% difference between predicted and the mean observed times). This result strongly consolidates our combination of different operators to point and select objects displayed inside the FoV.

5.6 Summary

We have consolidated our model through three different ecological tasks: 1) a pointing task outside the FoV, 2) a selection task outside the FoV and 3) a selection task inside the FoV. All observed times ranges from 1 to 4.8% of the predicted times computed with our model. Thus, our model appears to be an interesting approach to model the predicted times of fundamental tasks in mixed reality applications

6 CONCLUSION AND FUTURE WORK

In this paper we propose a new extension of KLM [6] for mixed reality applications modeling pointing and mid-air or device-supported validation tasks. First, we present a review of previous extensions of KLM and operators useful for these interactions. Next, we identified and described new operators required to model the pointing and validation tasks in mixed reality: Precise Head Point (**Php**), Coarse Head Point (**Chp**), Raise Hand (**Rh**), Air Tap (**At**) and Button Click (**Bc**). Through a set of user studies, we experimentally defined one or several unit times for each operator. Then, we consolidated our model through three user studies with ecological tasks combining multiple operators. The differences between the predicted times and the observed times are in the 20% error criterion from [33] which validates our KLM extension.

The use of KLM is based on the identification and combination of operators corresponding to atomic interaction steps. In the context of our studies, the gestures modelled by our operators are simple mechanical gestures that do not require any particular expertise. Moreover, the training phase and the embedded tutorial allowed the participants to become familiar with and master the required gestures. Consequently, it seems reasonable to consider that the times established in these works are independent of the degree of expertise of the user with mixed reality devices.

The use of the HoloLens v1 in our experiments offered a stable and robust Mixed Reality device with tracking integrated, which can favour the replicability of our work and application of our results given the widespread of this device. Regarding the HoloLens clicker, such type of buttons are available in other HMDs (e.g. Magic Leap One) and can be easily implemented using a physical button connected via Bluetooth to perform simple gestures. In general, the technologies and gestures recognized by head-mounted displays tend to evolve (hand-raycasting, direct touch). Given the rapid evolution of these technologies, further extensions of our work will be required. However, our paper presents a first systematic measurement of a set of fundamental operators for pointing and selecting digital elements.

To further support the exploration of this design space, several extensions can be envisioned. First, mixed reality devices allow the user to carry out other forms of interactions than those considered in this work (e.g. manipulation of 3D virtual objects). These interactions therefore require additional KLM operators to be appropriately modelled. However, these interactions can only be carried out after a search, pointing and selection task of a digital object. This is why we have focused in this first study on fundamental operators, a study which will require future extensions involving operators (e.g. Pinch,3D objects Grab and Drag, etc.) useful for more advanced tasks.

Finally, previous works on KLM extensions modeled smartphone-based interactions. Combining touchscreens with HMDs could be useful to overcome mixed reality mid-air interaction problems, such as fatigue [15, 16, 18], limited accuracy [38] and gesture discoverability [3]. To this end, it will also be interesting to extend our KLM model with operators for smartphone-based interactions (cf. Sect. 2.2) used in the context of Mixed Reality. This new model could be used to generate numerous interaction techniques.

ACKNOWLEDGMENTS

REFERENCES

[1] S. Al-Megren, J. Khabti, and H. S. Al-Khalifa. A Systematic Review of Modifications and Validation Methods for the Extension of the Keystroke-Level Model. *Advances in Human-Computer Interaction*, 2018:1–26, dec 2018. doi: 10.1155/2018/7528278

[2] C. Bailly, F. Leitner, and L. Nigay. Head-Controlled Menu in Mixed Reality with a HMD. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, and P. Zaphiris, eds., *Human-Computer Interaction – INTERACT 2019*, pp. 395–415. Springer International Publishing, Cham, 2019.

[3] T. Baudel and M. Beaudouin-Lafon. Charade: remote control of objects using free-hand gestures. *Communications of the ACM*, 36(7):28–35, jul 1993. doi: 10.1145/159544.159562

[4] R. Bernhaupt, D. Navarre, P. Palanque, and M. Winckler. Model-Based Evaluation: A New Way to Support Usability Evaluation of Multimodal Interactive Applications. pp. 96–119. 2008. doi: 10.1007/978-1-84628-941-5_5

[5] M. D. Byrne. *Cognitive Architecture*, pp. 97–117. L. Erlbaum Associates Inc., USA, 2002.

[6] S. K. Card, T. P. Moran, and A. Newell. The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 23(7):396–410, jul 1980. doi: 10.1145/358886.358895

[7] S. K. Card, T. P. Moran, and A. Newell. *The Psychology of Human-Computer Interaction*. CRC Press, may 1983. doi: 10.1201/9780203736166

[8] Y. Chen, K. Katsuragawa, and E. Lank. Understanding Viewport- and World-based Pointing with Everyday Smart Devices in Immersive Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13. ACM, New York, NY, USA, apr 2020. doi: 10.1145/3313831.3376592

[9] G. P. Chubb. SAINT, A Digital Simulation Language for the Study of Manned Systems. In *Manned Systems Design*, pp. 153–179. Springer US, Boston, MA, 1981. doi: 10.1007/978-1-4613-3306-7_6

[10] H. Dreyfuss Associates and S. B. Wilcox. *The Measure of Man and Woman: Human Factors in Design (Updated Edition)*. Wiley Online Library, 2001.

[11] K. El Batran and M. D. Dunlop. Enhancing KLM (keystroke-level model) to fit touch screen mobile devices. In *Proceedings of the 16th international conference on Human-computer interaction with mobile devices services - MobileHCI '14*, pp. 283–286. ACM Press, New York, New York, USA, 2014. doi: 10.1145/2628363.2628385

[12] B. Ens, D. Ahlström, and P. Irani. Moving Ahead with Peephole Pointing: Modelling Object Selection with Head-Worn Display Field of View. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, pp. 107–110. ACM, New York, NY, USA, oct 2016. doi: 10.1145/2983310.2985756

[13] O. Erazo and J. A. Pino. Predicting user performance time for hand gesture interfaces. *International Journal of Industrial Ergonomics*, 65:122–138, may 2018. doi: 10.1016/j.ergon.2017.07.010

[14] P. Green. Estimating Compliance with the 15-Second Rule for Driver-Interface Usability and Safety. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 43(18):987–991, sep 1999. doi: 10.1177/154193129904301809

[15] J. T. Hansberger, C. Peng, S. L. Mathis, V. Areyur Shanthakumar, S. C. Meacham, L. Cao, and V. R. Blakely. Dispelling the Gorilla Arm Syndrome: The Viability of Prolonged Gesture Interactions. pp. 505–520. 2017. doi: 10.1007/978-3-319-57987-0_41

[16] J. D. Hincapié-Ramos, X. Guo, P. Moghadasian, and P. Irani. Consumed endurance: A Metric to Quantify Arm Fatigue of Mid-air Interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1063–1072. ACM, New York, NY, USA, apr 2014. doi: 10.1145/2556288.2557130

[17] P. Holleis, F. Otto, H. Hussmann, and A. Schmidt. Keystroke-level model for advanced mobile phone interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '07*, pp. 1505–1514. ACM Press, New York, New York, USA, 2007. doi: 10.1145/1240624.1240851

[18] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani. Modeling Cumulative Arm Fatigue in Mid-Air Interaction based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3328–3339. ACM, New York, NY, USA, may 2017. doi: 10.1145/3025453.3025523

[19] Y. Jimenez and P. Morreale. Design and Evaluation of a Predictive Model for Smartphone Selection. pp. 376–384. 2013. doi: 10.1007/978-3-642-39253-5_41

[20] B. E. John and D. E. Kieras. Using GOMS for user interface design and evaluation. *ACM Transactions on Computer-Human Interaction*, 3(4):287–319, dec 1996. doi: 10.1145/235833.236050

[21] D. Kieras. Using the Keystroke-Level Model to Estimate Execution Times. Technical report, University of Michigan, Ann Arbor, MI, USA, 2001.

[22] D. Kieras. Model-based Evaluation. In A. Sears and J. A. Jacko, eds., *Human-Computer Interaction: Development Process*, chap. 15, p. 18. CRC Press, mar 2009. doi: 10.1201/9781420088892

[23] B. C. Kress. Digital optical elements and technologies (EDO19): applications to AR/VR/MR. In B. C. Kress and P. Schelkens, eds., *Digital Optical Technologies 2019*, p. 76. SPIE, jul 2019. doi: 10.1117/12.2544404

[24] K. R. Laughery. Micro SAINT — A Tool for Modeling Human Performance in Systems. In *Applications of Human Performance Models to System Design*, pp. 219–230. Springer US, Boston, MA, 1989. doi: 10.1007/978-1-4757-9244-7_15

[25] A. Lee, K. Song, H. B. Ryu, J. Kim, and G. Kwon. Fingerstroke time estimates for touchscreen-based mobile gaming interaction. *Human Movement Science*, 44:211–224, dec 2015. doi: 10.1016/j.humov.2015.09.003

[26] I. S. MacKenzie. Human-computer Interaction : An Empirical Research Perspective. In I. S. MacKenzie, ed., *Human-computer Interaction*, pp. i–iii. Elsevier, morgan kau ed., 2013. doi: 10.1016/B978-0-12-405865-1.00009-1

[27] D. Manes and P. Green. Evaluation of a Driver Interface: Effects of Control Type (Knob Versus Buttons) and Menu Structure (Depth Versus Breadth). Technical report, University of Michigan - Transportation Research Institute, Ann Arbor, MI, USA, 1997.

[28] D. McNeill. Hand and Mind: What Gestures Reveal about Thought. *Language*, 70(2):345, jun 1994. doi: 10.2307/415833

[29] Microsoft. Head-Gaze and Commit, 2021.

[30] Microsoft. MixedReality ToolKit, 2021.

[31] R. Myung. Keystroke-level analysis of Korean text entry methods on mobile phones. *International Journal of Human-Computer Studies*, 60(5-6):545–563, may 2004. doi: 10.1016/j.ijhcs.2003.10.002

[32] C. Nowakowski, Y. Utsui, and P. Green. Navigation System Destination Entry: The Effects of Driver Workload and Input Devices, And Implications for SAE Recommended Practice. Technical report, Universiy of Michigan - Transportaton Research Institute, Ann Arbor, USA, 2000.

[33] J. R. Olson and G. M. Olson. The Growth of Cognitive Modeling in Human-Computer Interaction Since GOMS. *Human–Computer Interaction*, 5(2-3):221–265, jun 1990. doi: 10.1080/07370024.1990.9667155

[34] P. G. Polson, C. Lewis, J. Rieman, and C. Wharton. Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of Man-Machine Studies*, 36(5):741–773, may 1992. doi: 10.1016/0020-7373(92)90039-N

[35] A. D. Rice and J. W. Lartigue. Touch-level model (TLM): evolving KLM-GOMS for touchscreen and mobile devices. In *Proceedings of the 2014 ACM Southeast Regional Conference on - ACM SE '14*, pp. 1–6. ACM Press, New York, New York, USA, 2014. doi: 10.1145/2638404.2638532

[36] S. Schneegaß, B. Pfleging, D. Kern, and A. Schmidt. Support for modeling interaction with automotive user interfaces. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '11*, p. 71. ACM Press, New York, New York, USA, 2011. doi: 10.1145/2381416.2381428

[37] A. Tilley and H. Dreyfuss Associates. The Measure of Man and Woman – Human Factors in Design. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 2(2):37–39, apr 1994. doi: 10.1177/106480469400200211

[38] E. Velloso, J. Turner, J. Alexander, A. Bulling, and H. Gellersen. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D

Manipulation. pp. 315–330. 2015. doi: 10.1007/978-3-319-22668-2_25