



Article

A Predictive Fingerstroke-Level Model for Smartwatch Interaction

Shiroq Al-Megren 

Information Technology Department, King Saud University, Riyadh 12371, Saudi Arabia; salmegren@ksu.edu.sa; Tel.: +966-11-805-7839

Received: 24 May 2018; Accepted: 25 June 2018; Published: 2 July 2018



Abstract: The keystroke-level model (KLM) is commonly used to predict the time it will take an expert user to accomplish a task without errors when using an interactive system. The KLM was initially intended to predict interactions in conventional set-ups, i.e., mouse and keyboard interactions. However, it has since been adapted to predict interactions with smartphones, in-vehicle information systems, and natural user interfaces. The simplicity of the KLM and its extensions, along with their resource- and time-saving capabilities, has driven their adoption. In recent years, the popularity of smartwatches has grown, introducing new design challenges due to the small touch screens and bimanual interactions involved, which make current extensions to the KLM unsuitable for modelling smartwatches. Therefore, it is necessary to study these interfaces and interactions. This paper reports on three studies performed to modify the original KLM and its extensions for smartwatch interaction. First, an observational study was conducted to characterise smartwatch interactions. Second, the unit times for the observed interactions were derived through another study, in which the times required to perform the relevant physical actions were measured. Finally, a third study was carried out to validate the model for interactions with the Apple Watch and Samsung Gear S3. The results show that the new model can accurately predict the performance of smartwatch users with a percentage error of 12.07%; a value that falls below the acceptable percentage dictated by the original KLM ~21%.

Keywords: keystroke-level model (KLM); fingerstroke-level model (FLM); smartwatch interaction; predictive model

1. Introduction

Wearable technologies refer to small smart devices that are attached or incorporated into body-worn accessories or pieces of clothing to provide convenient access to information [1]. Wearables are often light-weight, accessible when in motion, and facilitates control over data exchange and communication [2]. This recent trend is not merely perceived as a technological innovation, but it also holds a fashion component that can influence how and when they are worn [3]. The benefits of wearable technology can significantly impact societies and businesses due to their support of applications that promote self-care, activity recognition, and self-quantifying [2,4]. For instance, using wearables to track convulsive seizures could improve the quality of life of epileptic patients by altering caregivers [5]. Wearables also promote self-care by monitoring user activity and encouraging healthier behaviours and habits [6]. Another great benefit of wearable technology is its assistive capabilities within healthcare establishments, thus improving the success rates of medical procedures and safety of patients [7]. Within business, wearables can be used to train employees, improve customer satisfaction, and promote real-time access to information about goods and materials [8].

Among wearable technologies, the smartwatch stands out as being well understood due its familiarity to a traditional watch. In fact, the last few years have seen an increase in the popularity of commercial smartwatches. Recent statistics show that the sales of smartwatches have increased

from approximately 5 million units in 2014 to a staggering 75 million units in 2017 [9]. These numbers, which are projected to double in 2018, signify a level of smartwatch adoption that heralds smartwatches as the next dominant computing paradigm. As one of the latest developments in the evolution of information technology, the smartwatch offers its user a remarkable level of convenience, as it swiftly and discreetly delivers timely information with minimal interference or intrusion compared with smartphones and other mobile devices [10–12].

Much like a traditional watch, a smartwatch usually has a touchscreen and side buttons, through which user interaction is achieved. While these interaction modalities are similar to what is seen on a smartphone, the smartwatch provides a smaller touchscreen interaction space and requires bimanual interaction (i.e., the user's dominant hand is used to asymmetrically interact with the watch strapped to the non-dominant hand [13]). There has been a wealth of research addressing the potential capabilities and constraints of smartwatches. Several works have been carried out to gain an understanding of the use of smartwatches in daily activities (e.g., [12,14–16]), and even more work has investigated the mechanics of touch interactions on small devices (e.g., [17–19]). The interface design space of smartwatches has also received its share of research attention (e.g., [20–22]). The usability issues of smartwatches have recently been examined (e.g., [23,24]), as well as investigations of smartwatch acceptance (e.g., [25,26]). However, little effort has been made to assess smartwatch users' performance a priori to aid in the design of smartwatch applications.

Predictive models are commonly used in human–computer interaction (HCI) to analytically measure human performance and thus evaluate the usability of low-fidelity prototypes. The Goals, Operators, Methods, and Selection rules (GOMS) model encompasses a family of techniques that are used to describe the procedural knowledge that a user must have in order to operate a system, of which the keystroke-level model (KLM) is one of its simplest forms [27,28]. The KLM numerically predicts execution times for typical task scenarios in a conventional environment. The KLM was initially developed to model desktop systems with a mouse and keyboard as input devices, but it has since been extended to model new user interaction paradigms [27]. Several enhancements to the KLM have been applied to model touchscreen interactions with smartphones and tablets (e.g., [29,30]). The KLM has also been extended to in-vehicle information systems (IVISs), in which interaction often involves rotating knobs and pressing buttons (e.g., [31,32]). More recently, the KLM has been modified to address the expressive interactions used to engage with natural user interfaces (NUIs) [33]. Such modifications to the KLM have often proven valuable to the development process, as they reduce the cost of usability testing and help to identify problems early on.

To support future smartwatch development, this paper presents a pragmatic solution for designing and assessing smartwatch applications by building upon the well-established original KLM [27]. The main contributions of this paper are threefold. First, a revised KLM for smartwatch interaction, tentatively named Watch KLM, is introduced based on observations of participants' interactions with two types of smartwatches. The new model comprises fourteen operators that are either newly introduced, inherited, or retained from the original KLM [27]. Second, the unit time for each of these operators is revised to reflect the particularities of smartwatch interactions due to the smaller touchscreen and bimanual interaction mode. Third, the revised model's predictions are realistically validated by assessing its prediction performance in comparison with observed execution times. The results validate the model's ability to accurately assess the performance of smartwatch users with a small average error.

The remainder of this paper is organised as follows. Section 2 presents the background on the KLM, introducing the original model and its seminal related publications. Section 3 reviews modifications to the original KLM across various application domains. Section 4 describes the first experimental study, which was conducted to elicit revised operators for the modified KLM for smartwatch interaction. In Section 5, the second study is described, based on which the unit times for the revised operators are computed. The third study, conducted to validate the revised model for smartwatch interaction

to assess its efficacy at predicting task execution times, is reported in Section 6. Finally, Section 7 summarises and concludes the paper and briefly discusses future work.

2. Keystroke-Level Model

The KLM is considered to be one of the simpler variants of the GOMS family of techniques [28]. Unlike the other models in this family, the KLM calculates only the time it is expected to take an expert user to complete a task without errors in a conventional set-up. This calculation is based on the underlying assumption that the user employs a series of small and independent unit tasks; this assumption supports the segmentation of larger tasks into manageable units. In the KLM, the unit tasks are expressed in terms of a set of physical, mental, and system response operations. Each operator is identified by an alphabetical symbol and is assigned a unit value that is used in the calculation of an executable task (see Table 1). The KLM comprises six operations:

- The action of pressing a key, i.e., a keystroke, or pressing a button.
- The action of pointing to a target on a display with the mouse.
- The action of moving the hand between the keyboard and mouse or performing any fine hand adjustment on either device.
- The action of manually drawing a set of straight line segments within a constrained 0.56 cm grid using the mouse.
- A mental action operation to reflect the time it takes a user to mentally prepare for an action.
- The system response time to a user's action.

Table 1. KLM operators and unit times in seconds [27]. n_D is the number of straight line segments drawn with a total length of l_D .

Type	Operator	Time (s)
Physical	Keystroke (K)	0.20
	Point (P)	1.10
	Home (H)	0.40
	Draw (D)	$0.90n_D + 0.16l_D$
Mental	Mental action (M)	1.35
Response	System response ($R(t)$)	t

The mental action operator is unobservable but comprises a substantial fraction of a predicted execution time due to its representation of the time it takes a user to prepare to perform or think about performing an action. The placement of the mental operator is therefore governed by a set of heuristic rules that consider cognitive preparation. These rules are as follows:

1. Insert a mental operator in front of every keystroking operator. Also, place a mental operator in front of every pointing operator used to select a command.
2. Remove any mental operator that appears between two operators anticipated to appear next to each other.
3. Remove all mental operators except the first that belong to the same cognitive unit, where a cognitive unit is a premeditated chunk of cognitive activity.
4. Remove all mental operators that precede consecutive terminators.
5. Remove all mental operators that precede terminators of commands.

Once a series of physical and system response operations has been identified for a unit task and the placement of the mental operations has been determined, the KLM calculates the execution time of said task by summing the operators' unit times:

$$T_{execute} = T_K + T_P + T_H + T_D + T_M + T_R$$

$T_{operator}$ denotes the total time for a single operation; for example, $T_P = n_P t_P$, where n_P is the number of pointing actions and t_P is the duration of each pointing action.

To illustrate how the KLM's equation and rules can be applied to predict user performance, consider the following example of a user renaming a folder to 'test' on a desktop. The user homes the hand on the mouse, H ; points the mouse cursor at the object, P ; double-clicks on the folder icon to allow for renaming, KK ; homes hands on the keyboard, H ; keys new name 'test', $KKKK$, and presses Enter, K . The KLM model without M and R (assuming an instantaneous response from the system) is $HPKKHKKKKK$. Applying the heuristic rules for placing the M operators results in the final model $MHPKKHMKKKK$, where the first M is the time spent by the user searching for the folder on the computer display, and the second M is the time the user requires to mentally prepare for typing. Therefore:

$$T_{execute} = 2M + 2H + 1P + 7K = 2 \times 1.35 + 2 \times 0.40 + 1 \times 1.10 + 7 \times 0.20 = 6s$$

The performance of the KLM has been validated against observed values to determine its efficacy in predicting execution times. For this validation, the KLM was used to model typical tasks in various systems: executive subsystems and text and graphics editors. Expert users were asked to execute task scenarios to capture observed performances that were logged for comparison. The model's predictions were assessed in comparison with the observed values, and the root mean square percentage error (RMSPE) was calculated to be 21%. This level of accuracy is reported to be the best that can be expected from the KLM and is comparable to the values of 20–30% obtained with more elaborate models [28].

3. Related Work

Usability is defined as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use” [34]. There exist several methodologies that could be utilised to assess the usability of an interactive system or product. Usability tests can categorically be defined as expert-based (e.g., heuristic evaluation, cognitive walkthrough, and cognitive modelling) or user-based (e.g., lab-based testing with users) methods, each of which introduces benefits and limitations to the usability evaluation of an interactive system. Several advantages of wearable technology were previously recalled, and despite their potential as the next generation of core products in the IT industry, their commercial adoption has been slower than anticipated. To support the proliferation of wearables in the future, numerous studies have been conducted to identify the usability issues that arise from utilising wearable technology [2].

There have been numerous efforts that aimed to address and alleviate the usability issues of smartwatch usage (e.g., [23,24,35,36]). Usability evaluations were conducted to assess the usability of smartwatches in an academic setting, where a cheating paradigm was used as an example [36]. Quantitative and qualitative measures of effectiveness, efficiency, and usability were used to assess the ability of student's to cheat using a smartwatch. The results suggest the promotion of academic dishonesty when smartwatches are appropriated in academic settings. A design approach for wrist-worn wearables, which integrates micro-interaction and multi-dimensional graphical user interfaces, was proposed to offer guidelines to overcome the challenges of the interaction paradigmatic shift [23]. The usability of smartwatch operations was qualitatively studied to identify issues that impact information display, control, learnability, interoperability, and subjective preference [24]. The findings highlighted several limitations of wearable devices, such as its poor display of visually rich content and its variable interoperability. These findings lend themselves as design suggestions for user interface form factors that can improve the usability of smartwatches. Despite these efforts, none attempt to estimate ideal base times that can be utilised for quantifying usability with low-fidelity prototypes (e.g., KLM [27]).

The emergence of new technologies, such as wearables, introduces new challenges in the design and development of computer systems. Consequently, a need arises for the revision or

modification of conventional quality assessment models such as the KLM [27] for the purpose of usability quantification. Revised predictive models can help to evaluate human performance a priori and reduce the need for time- and resource-intensive human studies. The original KLM was developed to predict the performance of desktop users, i.e., in conventional set-ups, but it has continually been modified to model systems including various computing devices. The modifications made to the original KLM include the introduction of new operators, the adaptation of original operators or the adoption of operators from other KLM extensions, and the revision of heuristics or calculations. The rest of this section briefly reviews the literature on modified KLMs across different computing devices. Particular attention is paid to KLM extensions for touch-based smartphones due to the similarity of their interactions with those of smartwatches.

The earliest efforts made to extend the KLM focused on re-evaluating the original operators for applicability to new application domains or confirming its validity over time (e.g., [37–44]). The KLM was revised via a standardised four-dimensional methodology for the evaluation of text editors by introducing new operators and adapting those of the original model [37]. For measuring the performance of spreadsheet users, the KLM was first evaluated and later revised to consider cognition and extend its parameters [38]. Further modifications to the KLM were introduced to model skilled spreadsheet users' performance on hierarchical menus; for this purpose, the placement of the *mental action* operator was revised to account for skill [40]. For hypertext systems, the operators of the original KLM were extensively adapted to consider information retrieval with varying levels of expertise and system complexity [41]. For the assessment of user performance when using a history tool, the original KLM was extensively updated with ten new operators to address motor, cognitive, and perceptual operations that are relevant to the context of command specification in such tools [43]. More recently, almost 30 years after its conception, the original model was revised to explore the naturalistic behaviours through which users interact with today's computers [44]. Efforts have also been made to revise the KLM for the purpose of assessing the performance of users who are disabled (e.g., [39,42]). For instance, the KLM was extended for the performance assessment of users who are disabled when using a traditional set-up to access web content [42]. The operators in the revised model were expanded to include specific operations for shortcut access.

There has been a wealth of research on the modification of the KLM for the assessment of traditional and touchscreen IVIS designs prior to deployment (e.g., [31,32,45–50]). The value of the KLM in this domain lies in its ability to determine, simply and at a reduced cost, the time required to complete a task in order to assess the potential distraction it presents while driving. One of the earliest revisions of the KLM for traditional IVISs (i.e., with interaction via knobs and dials) involved adapting the model for the destination entry and retrieval tasks [49]. The original *keystroke* operation was extensively decomposed into finer, key-specific operations to increase the specificity of the model. To ensure that an IVIS is compliant with the 15 s rule for safe driving, a revised model was derived from the KLM that also considers adjustments for age [46]. The performance of a slightly revised KLM for IVIS designs was evaluated using an occlusion method that simulates driving behaviour (i.e., glancing between the road and the IVIS) for the purpose of determining domain-specific metrics and ensuring the observation of standards [31,47]. The performance of the modified model proved adequate, with significant correlations and low error rates. To support the creation of automotive interfaces, a predictive tangible prototyping tool (MI-AUI) was developed in which the unit times of operations are modified from the original KLM and used to estimate task completion time [48]. More recently, a revised version of the KLM [46] was further adapted for touch-based IVIS design by extending the considered set of operations to include gestures such as scrolling and dragging [50].

A large number of the studies concerned with extending the KLM to smartphone interactions consider its adaptation to text entry methods and predictive text (e.g., [51–55]). One of the earliest KLM extensions for text entry involved revising the original model for three text entry methods on a smartphone's 9-key keyboard [51]. The revised model was used to predict typing speeds considering various methods of input. Another modification to the KLM utilised the adapted

operations, Fitts' law, and a Chinese language model to predict user performance with two types of input on a smartphone [52]. The KLM was also extended to measure multi-finger touchscreen keystrokes for the purpose of representing 1Line, a text entry method for Chinese text input [53]. Revisions to the KLM for text entry have not only been limited to key input; another new model investigated the feasibility of a speech-based interface design for text messages [56]. The evaluations conducted to determine the effectiveness of the design choices indicated the usefulness of speech input. Modifications to the original KLM for smartphone interaction beyond the umbrella of text entry have also been considered. One of the earliest models to comprehensively consider key-based smartphones was the Mobile KLM, in which new operations were introduced to account for interactions with key-based smartphones [29]. This model considered interactions on and with the smartphone, e.g., key presses and gestures. The unit times were also analysed for each of the new and adapted operations. This model was later refined to consider near-field communication (NFC) and dynamic interfaces [57].

Over the past decade, touch-based smartphones have almost entirely replaced traditional smartphones, thus necessitating revised extensions to the KLM to properly assess this new paradigm of human interaction. While smartphones are dissimilar in size to smartwatches, the touch-based interactions with both types of devices are somewhat similar. One of the earliest proposed variants of the KLM for touch-based smartphone interactions considers stylus input [58]. In the new model, the generality of the *mental action* operations in the original KLM is reassessed, and the corresponding operator is split into five new mental operators for task initialisation, decision making, information retrieval, finding information and verifying input. The *homing* operator distinguishes between homing a stylus to a certain location and homing a finger. Other physical operations include tapping, picking up the stylus, opening a hidden keyboard, rotating the smartphone, pressing a side key, and plugging/unplugging a device to/from the smartphone. The proposed model also introduces the concept of an operator block (OB) to indicate a sequence of operations that are likely to be used together with high repeatability.

The touch-level model (TLM) is a revised version of the original KLM that incorporates interactions with modern touch-based devices, such as smartphones [59]. The new model retains several operators from the original KLM, namely, *keystroking*, *homing*, *mental action*, and *response time*, as they remain relevant for touch input. The rest of the KLM operators were either considered inapplicable to the context or retained within other operations. Several operators were also inherited or adapted from the Mobile KLM, including initial action and distraction [29]. Numerous new operators were also proposed to allow the new model to account for touchscreen interactions and gestures. These operators include gesturing, pinching, zooming, tapping, swiping, tilting, rotating, and dragging. While the model has the potential to be used for benchmarking touchscreen interactions on a smartphones, the new operators currently have no baseline unit values and are yet to be validated. Three of the proposed operations in TLM (tap, swipe, and zoom) were also reflected in a new model that modifies KLM for touch-based tablets and smartphones [60]. Fitts' law, a descriptive model that predicts that time it will take for a user to point to a target, was utilised in combination with this model to determine the unit execution times for each of the proposed operations using custom prototypes. Unit times were suggested for short swipes (0.07 s), zooming (0.2 s), and tapping on an icon from a home position (0.08 s). More accurate estimates are possible with the proposed predictive equations.

The fingerstroke-level model (FLM) is a revised version of the KLM for the assessment of mobile-based game applications on touch-based smartphones [61]. The proposed model adapts several operators from the original model for application to touch-based interactions, including tapping (adapted from *keystroking*), *pointing*, and dragging (adapted from *drawing*). The FLM uses the original *mental action* and *response time* operators from the original KLM [27] and introduces a new flicking operator. For each of the new and adapted operators, unit times were determined through an experimental study to ensure accuracy with respect to touch interactions. Regression models were also devised for four of the operators (tapping, pointing, dragging, and flicking). The validity of the model for estimating execution times for mobile gaming was verified in an experimental scenario,

in which its predictions were more accurate than those of the original KLM. An extension to the FLM, the Blind FLM, was later proposed for actions performed by smartphone users who are blind [62]. This model was similarly validated, resulting in a low root mean square error of 2.36.

Of the four models adapted from the KLM for touch-based interaction, the FLM [61] and the work of El Batran et al. [60] include unit times for each of the proposed operators. The FLM was the only model validated in experimental studies, in which it was found to be effective for modelling mobile gaming applications. Table 2 summarises the operators of the revised KLM models for touch-based interaction in relation to the original KLM. The FLM, the TLM, and El Batran's model share *tapping*, *mental action*, and *dragging* with the original KLM. *Swiping* is similarly shared across these three adapted models but not with the original KLM. *Tapping* appears in all four models and was arguably adapted from the original *keystroking* operator. While the TLM comprehensively covers all possible touchscreen interactions, it lacks unit times and remains to be validated. For this reason, the revised KLM for smartwatch interaction will be discussed in relation to the FLM in the following sections.

Table 2. Revised versions of the KLM for touch-based smartphones. The models are presented in relation to the original KLM [27]. It should be noted that Li et al. [58] revised model considers the use of a stylus for interaction.

KLM	Li et al. + [58]	TLM [59]	El Batran et al. [60]	FLM [61]
Keystroke	Tap Press a key on the side Press and hold	Tap Keystroke	Tap	Tap
Point	×	×	×	Point
Home	Home stylus Home finger	Home	×	×
Draw	×	Drag	Swipe _{long}	Drag
Mental action	Mentally initiate a task Mentally decide or choose Mentally retrieve Mentally find Mentally verify	Mental action	×	Mental action
System response	×	×	×	System response
<i>New Operations</i>				
	Flip or slip keyboard	×	×	×
	Pick up stylus	×	×	×
	Rotate phone	Tilt	×	×
	Plug/unplug	×	×	×
	×	Swipe	Swipe _{short}	Flick
	×	Rotate	×	×
	×	Initial action	×	×
	×	Gesture	×	×
	×	Distraction	×	×
	×	Pinch	×	×
	×	Zoom	Zoom	×

4. Study 1: Operator Extraction for Smartwatch Interaction

An observational study was conducted for the purpose of identifying the range of interactions or gestures that people perform when interacting with a smartwatch. Therefore, the main objective of this first study was to determine the physical operators that can potentially compose the execution of a task on a smartwatch. Prior to the study, a web-based questionnaire was distributed to identify the smartwatches and first- or third-party applications commonly used in the community. The questionnaire received 73 respondents, of whom 48 were female and 25 were male. The respondents had an average age of 28.4 years, with a standard deviation of 6.2 years. The majority of the participants (64%) were frequent users of smartwatches, with an average of 9.2 h of daily utilisation. The participants declared

over 22.7 h of average weekly use. The Apple Watch (watchOS) and Samsung Gear (Tizen OS) were the two most regularly used smartwatches among 78% of the participants. Other smartwatches mentioned by the other participants included Pebble, Fitbit, and LG Watch. Approximately 8% of the respondents reported that they were familiar with more than one smartwatch. First-party applications, with a 79% utilisation rate, were more commonly utilised than third-party applications. Message notification applications were the most commonly utilised first-party application type, with utilisation rate of 86%, followed by built-in activity and health applications and, less frequently, by virtual assistants (e.g., Siri or S Voice). Notable classes of third-party applications included social media (e.g., WhatsApp and Twitter), health and exercise (e.g., Nike Plus, Runtastic, and Endomondo), and news (e.g., Flipboard and BBC). The analysed data were used to guide the selection of smartwatches and applications to be considered throughout this paper.

4.1. Methodology

4.1.1. Participants

Twenty-two participants (16 female and 6 male) were recruited for the observation of smartwatch interactions. The participants were recruited via social media, advertisements on university campuses, and local activity websites. The average age of the participants was 24.2 years, with a standard deviation of 3.4 years. All participants used their smartwatches regularly in their daily lives. Twenty of the participants were right-handed, two were left-handed, and all wore their smartwatches on their non-dominant hand. All participants owned smartwatches; 17 participants had an Apple Watch, 4 had a Samsung Gear, and 1 participant owned both and ended up being observed using the Samsung Gear smartwatch. Table 3 further specifies the distribution of the participants in terms of the watch type and screen dimensions of the smartwatch used. This variety of smartwatches was welcomed to encompass all possible interactions with the predominantly used smartwatches in the community. All participants used their smartwatches on a daily basis, with an average daily use of 8.4 h. All participation was voluntary, and each participant was compensated with a gift voucher for their time.

Table 3. Participant distribution across smartwatches in terms of smartwatch type and screen dimensions.

Smartwatch	Participants	Screen Dimensions
Apple Watch Series 1	11	38.6 × 33.3 mm
Apple Watch Series 2	3	38.6 × 33.3 mm
Apple Watch Series 2	3	42.5 × 36.4 mm
Samsung Gear S2	2	42.3 × 49.8 mm
Samsung Gear S3	3	42.9 × 44.6 mm

4.1.2. Materials

Two main types of smartwatches were used by the participants in this study: Apple Watch and Samsung Gear. These two smartwatch types were reported to be the most commonly used in the community based on the on-line questionnaire. There were three variants of the Apple Watch and two of the Samsung Gear in terms of the style and screen dimensions (see Table 3). The Apple Watch has a rectangular watch face with two buttons on the side: crown and side button. The Samsung Gear has a circular watch face and similarly has two side buttons and a rotating bezel. Two cameras were used to record the participants' interactions with their smartwatches. One camera (SJ4000 Sports Action DV 12MP) was strapped to each participant's shoulder to provide a rough approximation of the participant's view of the smartwatch [15]. The second camera (Canon PowerShot SX60 16.1MP) was portable to allow the observer to monitor the behaviour of the participants as they moved about the study room. Since the goal of this first study was to analyse the participants' smartwatch interactions while preserving a degree of ecological validity, the procedure and tasks in this first study were uncontrolled.

4.1.3. Procedure

The observation sessions were held in a quiet room and lasted approximately 19 min. Each session was held with one participant at a time. Each session began with welcoming the participant to the study. The participant was then introduced to the general research objective. An information sheet was presented to the participant, which provided further details regarding the purpose of the study and the information to be collected. The participant was then prompted for any questions regarding the study. A consent form was later presented to be signed. The shoulder camera was strapped to the participant's shoulder on the same side as the non-dominant hand. The placement of the camera was adjusted to ensure a clear view of the smartwatch by asking the participant to perform a series of interactions. The participant was then asked to use the watch as normal, whether standing, sitting, or walking around the study room. All sessions were video recorded. The session was finally closed by thanking the participant and providing compensation for participating.

4.1.4. Data Analysis

The analysis of the video data began with reviewing the videos and extracting clips that involved any form of smartwatch interaction from the two captured views (shoulder camera and portable camera). The video recordings were then coded using the Behavioral Observation Research Interactive Software (BORIS) [63]. After this, the physical interactions were sorted into three main categories:

- Touchscreen interactions: these are all interactions performed on the touch-sensitive watch face.
- Device interactions: this category includes all interactions performed on the side buttons of the smartwatch or bezel. Both observed smartwatch types had two buttons on the side, and the Samsung Gear had a rotating bezel.
- Gestures: this category includes all interactions performed by the participants that do not directly involve touch or button actions with the smartwatch. The category also involves actions around the smartwatch.

4.2. Findings

The average session time was 18.27 min, with a standard deviation of 2.64 min. On average, each video contained 63.59 actions, with a standard deviations of 7.2 actions. The total number of actions analysed from all participants was 1399. Of those actions, 67% were touchscreen interactions, 29% were device interactions, and 4% were gestures (see Figure 1). The majority of the actions were observed from the participants for both smartwatches, the Apple Watch and the Samsung Gear. Table 4 summarises the actions observed in relation to the original operators of the KLM [27] and the FLM [61]. The revised KLM for smartwatch interaction is kept as general as possible to ensure that the operators can provide accurate estimates when used to model most other types of smartwatches. The rest of this section describes the differences and similarities between the revised KLM for smartwatches, the original KLM [27] (see Section 2), and the FLM (see Table 2) [61].

4.2.1. Adapted Operators

Tap This action was the most commonly observed action in the study. Tapping on the watch face is used to initiate an action or make a selection from a set of options. This action is performed using one or two fingers and most commonly using either the index finger of the dominant hand or the index and middle fingers, in the case of two-finger tapping. Several participants were observed to use their thumbs for tapping operations. For instance, to open an app on either the Apple Watch or the Samsung Gear, a participant would tap on the application's icon to open it. The hand postures when tapping on the screen varied among the participants. In previous work, these postures have been classified as no-support (NS), thumb-support (TS), index-finger-support (IS), and thumb-and-middle-finger-support (TMS) postures [15,64]. NS and TS postures were the most

commonly observed in this study. In relation to the original KLM [27] and the FLM [61], tapping corresponds well to the keystroking and tapping operators, respectively.

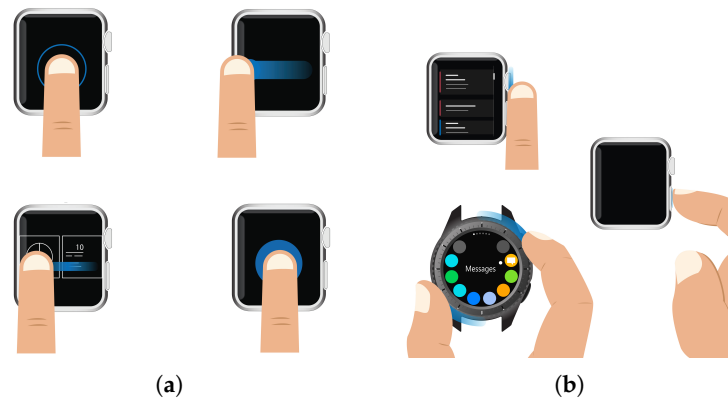


Figure 1. Some of the physical operations performed on the Apple Watch and Samsung Gear. (a) Touchscreen interactions [left to right, top to bottom]: *tap*, *drag*, *swipe*, and *tap-and-hold*; (b) Button and bezel interactions [top to bottom]: *turn button*, *press button*, and *turn bezel*.

Table 4. Observed operators for the smartwatch-extended KLM in relation to the original KLM [27] and the FLM [61].

Operator	Description	KLM Comparison	FLM Comparison
<i>Touchscreen Interactions</i>			
Tap	A tap on some area of the touchscreen with one or two fingers to initiate an action or make a selection.	Keystroke	Tap
Double tap	A double tap on some area of the touchscreen with one or two fingers to initiate an action.	Keystroke $\times 2$	Tap $\times 2$
Tap-and-hold	A long, firm press on the touchscreen to activate an action or command or to change modes.	Not applicable	Not applicable
Drag	A gesture that involves pressing on an on-screen target and then moving from one target or view to another.	Draw	Drag
Swipe	A gesture that is faster than a drag and typically has no on-screen target.	Not applicable	Flick
<i>Device Interactions</i>			
Press	A press of a button on the side of the smartwatch.	Keystroke	Not applicable
Double press	A double press on a side button.	Keystroke $\times 2$	Not applicable
Press-and-hold	A press of a side button followed by a hold.	Not applicable	Not applicable
Turn	A turn or rotation of a dial on the side of the smartwatch.	Not applicable	Not applicable
	A turn or rotation of a dial around the smartwatch face.	Not applicable	Not applicable
<i>Gestures</i>			
Initial action	A gesture to initiate an action with the smartwatch. It requires the user's arm to move such that the watch face is facing upwards.	Not applicable	Not applicable
Home	The homing of the hand from the <i>initial action</i> to the smartwatch.	Home	Not applicable
	The homing of the hand and fingers from the smartwatch face to a button or vice versa.	Home	Not applicable
Gesture	Any other gesture supported by the smartwatch device.	Not applicable	Not applicable
<i>Mental and System Operations</i>			
Mental action	The mental time required to prepare for an action.	Mental action	Mental action
System response	The time required for the smartwatch to react to an action.	System response	System response

Double Tap Several smartwatch interactions require the user to double tap on the touchscreen using the index finger, or, less frequently, the thumb, to initiate an action. This action is supported by both the Apple Watch and Samsung Gear smartwatches. On the Samsung Gear, double tapping with a single finger was used by one of the participants to zoom in on a map. In the case of the Apple Watch, zooming was achieved via two-finger double tap. Similar to tapping, this action is analogous to performing two consecutive keystroke actions in the original KLM [27] or two consecutive taps in the FLM [61] using one or two fingers.

Drag This action is performed by dragging an on-screen target from one point to another. The act of dragging was observed to be either bounded or unbounded. A bounded action, e.g., turning on the reserve battery on the Apple Watch, requires the vertical or horizontal movement of a target within a control element. By contrast, an unbounded drag has two defined end points, namely, the starting position and the target, but the movement is not constrained by any control element. For instance, on the Samsung Gear, the icon for an application can be moved from one position to another without any constraints. In relation to the original KLM, this action is somewhat similar to the *drawing* operation [27]. The dragging action is one of the operations considered in the FLM, where it is defined as the act of moving an object on the touch screen [61]. The hand posture typically associated with this action was the TS posture.

Swipe The second most commonly observed touchscreen interaction was a swiping action, which is a faster gesture than a drag and is typically not associated with an on-screen target. The participants were typically observed to use this gesture to navigate options or panels on the Apple Watch and Samsung Gear. Unlike the drag action, the swiping action was most often observed to be associated with the NS or TS hand posture. In relation to the FLM, this action is similar to the *flicking* operation, which is defined as dragging as quickly and over as short a distance as possible [61]. Swipe operations are not analogous to any of the original operators of the KLM [27].

Press The action of clicking on a side button on a smartwatch is referred to here as a press. Of the 29% of the observed smartwatch interactions that were physical device interactions, pressing a side button was the most common. The participants were commonly observed to perform this action with the index finger of the dominant hand while using the thumb to support the free hand (akin to what was observed for tapping interactions [15,64]). For example, a participant with a Samsung Gear S3 watch regularly used this action to backtrack to a previous state. In relation to the original KLM, this action corresponds very well to the *keystroke* operator [27]. Since the FLM considers only touch interaction, none of its operators is analogous to the *press* action [61].

Double Press Several of the observed smartwatch interactions required the user to double press on a side button to perform an action. This action was not as common as the press operation and was only observed from some of the participants. This is expected, since the operations associated with this action are limited. On the Apple Watch, one participant was observed to double press to view Apple Pay. Similar to the press operation, a double press is analogous to performing two consecutive *keystroke* actions in the original KLM [27].

Home A homing gesture is one of two types of actions that constitute the act of homing. The first often followed the initial action gesture, in which a participant would bring the non-dominant arm with the smartwatch closer to perform an action, and was performed to home the dominant hand to the smartwatch. The second class of homing actions encompasses the micro-actions performed to home the hand and fingers from the watch face to the buttons or vice versa. Actions of the latter type were observed more regularly as the participants interacted with their smartwatches. Actions of the former type were observed not only following initial action gestures but also when bringing the arms closer together for interaction. For instance, one of the participants was sitting down with both arms on the desk and would regularly bring the dominant hand close to the watch when taking breaks from navigating the smartwatch. Both gestures are analogous to the *homing* operator in the original KLM [27]. The *homing* operator was considered to be of less use in the FLM within the context of mobile gaming [61].

4.2.2. New Operators

Tap and Hold The tap-and-hold action is often referred to as a force touch on the Apple Watch and a press on the Samsung Gear. This operation is initiated by a tap that is followed by a variable hold time that depends on the action being activated. Tap-and-hold actions were observed to be performed a few times using the index finger or thumb by the majority of the participants. For instance, on both the Samsung Gear and the Apple Watch, the participants tapped and held on the application view display to rearrange applications. Similar to what has been reported above for the *tap* and *double tap* operations, NS and TS postures were observed for most participants. The tap-and-hold action is not applicable in either the original KLM [27] or the FLM [61].

Press and Hold This action constitutes two discrete operations: a press operation (as previously described) followed by a variable hold time. The hold time depends on the action being initiated. The measured frequency of this action was relatively low, with only nine observed instances. An example of the press-and-hold action was observed when one of the participants performed this operation to use Siri on the Apple Watch. Similar to the main posture identified for the press action, the support of the thumb was often observed when performing the press-and-hold action. This operation has no relation to the operators of the original KLM [27] or the FLM [61].

Turn On both the Apple Watch and the Samsung Gear, the turning operation entails the rotation of a dial mechanism either clockwise or counter-clockwise to navigate options. In the case of the Apple Watch, a turning operation is performed by rotating one of the side buttons (the digital crown button). By contrast, the Samsung Gear supports the rotation of the bezel surrounding the watch face. Both types of turning operations were observed in this study. However, the frequency of bezel rotations was much higher than that of crown rotations. The most commonly observed hand posture for a bezel turn involved the placement of the index finger and thumb on opposite sides. Only two participants were witnessed to rotate the bezel with only the index finger. Crown rotation was similarly observed using one or two fingers. The turning operation was not previously considered in either the original KLM [27] or the FLM [61]. However, a turning operator has previously been utilised to represent operations on physical dials in IVISs (e.g., [32,48]).

Initial action In the original KLM, it is assumed that the user is positioned in front of the desktop with both hands on the keyboard and mouse [27]. The same assumption is carried over to the FLM, where it is assumed that the user is holding the smartphone horizontally with one hand and has positioned the other hand next to the phone [61]. The smartwatch scenario introduces a different setting in which the observed participants must perform an initiating action prior to performing a task. This initial action was similar for both the Samsung Gear and the Apple Watch: the participant was required to rotate his/her arm and bring it closer at chest level to essentially wake up the device from sleep mode before proceeding with an action. This initial action was similar to the of the revised Mobile KLM for mobile phone interaction [29].

Gesture A silencing gesture is supported by both the Samsung Gear and the Apple Watch. In such a gesture, the smartwatch user covers the watch face to silence the device or put it to sleep. This action was only observed to be performed by two of the participants when a phone call was received and the dominant hand was used to cover the watch face. Although no other gestures were witnessed or supported for interaction with the smartwatch devices included in this study, potential extensions to these devices based on their already built-in capabilities (e.g., accelerometer and gyroscope) necessitate their inclusion in the model. While gestures are not considered in either the original KLM [27] or the FLM [61], this *gesture* operator is similar to that of the revised Mobile KLM for mobile phone interaction [29].

4.2.3. Retained Operators

Mental action The mental act operation refers to the fact that experienced users will often pause for approximately one second while performing routine actions in order to remember, find, or process something prior to executing a physical action [65]. This operator is retained in its original form from

the original KLM [27]. However, since new operators have been observed and introduced into the revised model, the original KLM mental heuristics are modified as follows:

1. Insert a mental operator in front of every *tap*, *press*, *drag*, *swipe*, *turn*, or *gesture* operator. Also, place a mental operator in front of every compound operator, including *double tap*, *tap-and-hold*, *press-and-hold*, and *double press*.
2. Remove all mental operators that appear between two operators anticipated to appear next to each other.
3. Remove all mental operators except the first that belong to the same cognitive unit, where a cognitive unit is a premeditated chunk of cognitive activity.
4. Remove all mental operators that precede redundant terminators.
5. Emphasise the number of mental operators more than the placement of those operators [29].

System response This operator is adapted from the original KLM to represent the time it takes for the system to respond to a user's action [27]. This time is system dependent and is arguably irrelevant given the relevant technological advancements. According to the observations made in this study, a smartwatch's response to a user action is typically instantaneous, and thus, the unit time value of this operator is arguably negligible. Nevertheless, this operator is maintained in the revised KLM for smartwatch interaction to account for different devices and software.

5. Study 2: Operators' Unit Times

For the revised KLM for smartwatch interaction to be used in practice to predict the time it will take a user to complete a task, the unit time for each of the identified operators must be determined. Therefore, the main objective of this study was to determine the time estimate for each physical operator identified in the previous section for smartwatch interaction.

5.1. Methodology

5.1.1. Participants

To determine the operators' unit times for the modified smartwatch KLM, thirty participants (16 female and 14 male) were recruited for this experiment. The study was advertised via social media, board announcements around campus, and local activity websites (e.g., local WhatsApp groups). The participants had an average age of 25.9 years, with a standard deviation of 6.05 years. All participants owned an Apple Watch, which they regularly used in their daily lives, with an average daily use of 6.1 h and 1.9 years of overall experience. The recruited participants were all right-handed and wore their smartwatches on the left hand (i.e., the non-dominant hand). Of the thirty participants, only two had taken part in the previous study. Participation was voluntary, and each participant was compensated with a gift voucher for participating in the experiment.

5.1.2. Materials

As seen from the questionnaire issued prior to the first study (see Section 4), the respondents attested to the dominance of the Apple Watch as the most common smartwatch. Of the 73 respondents, 69% owned an Apple Watch and used it on a regular basis. For this reason, the Apple Watch was chosen for use in this study over the other reported smartwatches (e.g., Samsung Gear, Pebble, Fitbit, and LG Watch). While each participant owned his/her own Apple Watch, each participant used the same smartwatch in this experiment to ensure that the initial states were uniform across all participants. The smartwatch used in this study was an Apple Watch Series 1 with screen dimensions of 38.6×33.33 mm. The participants' interactions were captured using video cameras and a custom application developed for the Apple Watch. Similar to study 1, two cameras (an SJ4000 Sports Action DV 12MP and a Canon PowerShot SX60 16.1MP) were used to record the interactions; the SJ4000 was strapped to the participant's shoulder, and the Canon PowerShot was positioned with a direct

view of the participant and smartwatch. A custom application was also used to isolate interactions and eliminate ambiguity. Of the 14 operators identified for smartwatch interactions (see Table 4), the first 12 were measured in this study (see Table 4); the *mental action* and *system response* operations are unchanged from the original KLM [27]. In total, there were 22 input tasks that spanned the 12 operations to be measured. The basic inputs for each of the 12 operators were organised into 8 groups of interactions as follows:

- *Tap* group: This group include the *tap*, *double tap*, and *tap-and-hold* operations. In the custom application, a circular target size of 75 px was set, following the Apple Watch human interface guidelines [66]. *Tap-and-hold* actions were captured without a target, as is typically the case when this action is executed in a natural setting.
- *Drag*: The *drag* operation was captured in five directions: left, right, up, down, and diagonal. In the custom application, these actions were captured with a bounded control for each of the four directions.
- *Swipe*: Similarly to *drag*, this operation was logged in five directions: left, right, up, down, and diagonal. However, *swipe* actions are unbounded, and in the custom application, they were captured for each of the four directions by asking the participants to navigate options vertically and horizontally.
- *Button press* group: This group of operations includes the *press*, *double press*, and *press-and-hold* operations that are performed using the side buttons of the Apple Watch.
- *Button turn*: Similarly to the *swipe* action, this operation was captured in the custom application by asking the participants to navigate options vertically by rotating the digital crown upwards and downwards in approximately 18-degree increments.
- *Initial action*: This action was captured on the video feed when the study moderator asked the participant to relax and then bring up the non-dominant arm for smartwatch interaction.
- *Home* group: This group includes micro-homing actions between the side buttons and the touch screen as well as the act of homing the dominant hand after the *initial action* of bringing the non-dominant arm closer.
- *Silencing gesture*: To induce the silencing *gesture*, the moderator initiated a phone call with the study's Apple Watch. The time it took to perform this action was captured via the video feed.

5.1.3. Procedure

The procedure took approximately 22.27 min and was held in a quiet room. Each session began with welcoming the participant to the study, followed by an introduction to the overall objective of the study. An information sheet was presented to the participants informing them of the particularities of the study and prompting any questions. The consent form was then signed by all participants. The study's smartwatch was placed on the participant's non-dominant hand, and the participant was asked to adjust its placement on the wrist to ensure user comfort. The shoulder camera was strapped to the participant's shoulder on the non-dominant side, which, in this case, was the left shoulder for all participants. Adjustments were made to the placement to ensure comfort and a clear view of the smartwatch. The other camera was adjusted to have a view of the participant standing or sitting and his/her interactions with the smartwatch. A brief training session preceded the input tasks, in which the participant was asked to perform all 12 operations using the custom application. A within-subject design approach was adopted for this experiment, with all operations performed by each participant. The input tasks were randomly ordered for each participant. Each participant was asked to complete the 22 input tasks covering the 12 operations, with three trials for each of the input tasks. The participants were not asked to use any specific hand posture to ensure that their behaviour was as natural as possible. Each session was finally closed by thanking the participant and providing compensation for participating.

5.1.4. Data Analysis

The majority of the unit times were captured automatically using the custom application. These included the times for the following operators: *tap*, *double tap*, *tap-and-hold*, *drag*, *swipe*, *press*, *press-and-hold*, *double press*, and *button turn*. In the custom application, the unit time for each operation was computed from the moment of initiation (i.e., the placement of the finger on the touchscreen or button) to the completion of the input task. The unit times for the remaining interactions, categorised as gestures in the previous study, were solely extracted from the video feed. The video data were analysed by reviewing the feed and separating the extracted clips into the expected 12 operations performed by the participants with the Apple Watch. These actions were the *initial action*, *homing*, and the silencing *gesture*. Manual coding of the start and end times of each gesture action was completed using BORIS [63].

5.2. Findings

The findings from this study are the average time taken by a smartwatch user to complete each of the identified smartwatch operations (see Figure 1 and Table 4) in the series of input tasks. All 1980 trials ($22 \times 3 \times 30$) were successfully completed. This excludes the 22 trials that were performed for training purposes. Table 5 summarises the time estimates for the 12 physical operations for smartwatch interaction. The differences among the *swipe* and *drag* operations in different directions were minimal, and thus, a single value for each operation is presented in the table. The unit times for *double tap* and *double press* confirm a carry-over effect from the first *tap* or *press*, respectively. The unit time for the mental action operator is adopted from the original KLM, with a fixed value of 1.35 s [27]. Most previous studies have similarly used the original mental action value (e.g., [29,31,42,61]), although smaller values have also been reported for specialised applications (e.g., [54,67,68]).

Table 5. Proposed average times and standard deviations for all operators identified in study 1, as reported in Section 4. The *mental action* and *system response* times are retained from the original KLM [27].

Operator	Average Time (s)	Standard Deviation (s)
Tap	0.28	0.16
Double tap	0.43	0.10
Tap and hold	0.78	0.09
Drag	0.55	0.21
Swipe	0.18	0.04
Press	0.47	0.05
Press and hold	0.95	0.15
Double press	0.90	0.19
Turn	0.47	0.16
Initial action	0.82	0.11
Home	0.57	0.09
Home _{micro}	0.34	0.11
Gesture	0.75	0.17
<i>Retained Operator Values</i>		
Mental action	1.35	
System response	<i>variable</i>	

6. Study 3: Model Validation

The first two studies extracted physical operators and determined a unit time estimate for each of those operators, respectively. The third study's objective was to validate the realistic predictions obtained with these operators by computing the accuracy of the new model by comparing the observed execution times with the predicted times.

6.1. Methodology

6.1.1. Participants

For this study, forty individuals (23 female and 17 male) were recruited to assess the predicted execution times against the observed times. The participants were divided into two groups based on the smartwatch they owned: Apple Watch group and Samsung Gear group, with twenty participant in each group. The Apple Watch group had an average age of 23.5 years, with a standard deviation of 3.6 years. The Samsung Gear participants had an average age of 26.3 years, with a standard deviation of 3.9 years. On average, the participants had an average of 1.8 and 2 years of experience using a smartwatch in the Apple Watch and Samsung Gear groups, respectively. All participants reported that they had been using their smartwatches daily, with an average daily use of 6.5 h. All recruited individuals were right-handed and wore their smartwatches on their non-dominant hands. As in the two previous studies, recruitment was carried out via social media, advertisements around campus, and local activity websites (e.g., WhatsApp groups). None of the participants in this study had taken part in any of the earlier studies. Participation was voluntary, and the participants were compensated with a gift voucher for their time.

6.1.2. Materials

While the recruited participants owned their own smartwatches, for this study, a test Apple Watch Series 3 (38.6 × 33.3 mm) and a test Samsung Gear S3 (42.9 × 44.6 mm) were used. This ensured the uniformity of the task steps and initial states. An SJ4000 Sports Action DV 12MP camera was used to log the observed execution times. The camera was strapped to the participant's shoulder to ensure a clear view of the participant's interactions with the smartwatch. Five experimental tasks were examined to validate the accuracy of the revised KLM for smartwatch interactions. The tasks were different for the two smartwatches, either in nature or in the series of actions. This was not a concern since the purpose of the study was to assess the accuracy of prediction for a given task. The majority of the tasks for both smartwatches involved first-party applications because of their popularity among the respondents to the initial questionnaire (see Section 4). The tasks were also designed to incorporate all operators that were identified for the revised KLM for smartwatch interaction. The five experimental tasks and their associated actions were as follows for the Apple Watch:

1. Change watch face—In this task, the participants were asked to change the watch face to the third option on the right. This task involved the *initial action*, *home*, *mental action*, *tap-and-hold*, *swipe*, and *tap* operations.
2. Set alarm—The participants were asked to set an alarm for 4 h from the current time. This task involved the *initial action*, *home*, *mental action*, *press*, *drag*, *tap*, and *turn* operations.
3. Open map and zoom × 2—In this task, the participants were asked to open the local map application and zoom × 2 on the current location. The operations involved in this task included the *initial action*, *home*, *mental action*, *tap*, and *double tap* operations.
4. Turn on Siri, then put to sleep—In this task, the participants were asked to initiate interaction with Siri and then put the smartwatch to sleep. This included the following operations: *initial action*, *home*, *mental act*, *press-and-hold*, and *gesture*.
5. Initiate a workout—For this task, the participants were asked to initiate a workout by selecting the 3rd exercise option. The following operations were utilised in this task: *initial action*, *home*, *mental action*, *double press*, *swipe*, and *tap*.

Several tasks were shared between the Apple Watch and the Samsung Gear S3. Nevertheless, the set of utilised actions for a given task often differed. The five experimental tasks for the Samsung Gear S3 were as follows:

1. Change watch face—For this task, the participants were asked to change the watch face to the third option from the right. The operations that were involved in this task were the *initial action*, *home*, *mental action*, *tap-and-hold*, *swipe*, and *tap* operations.
2. Set alarm—The participants were asked to set an alarm for 4 h from the current time. This task involved the *initial action*, *home*, *mental action*, *press*, *drag*, *tap*, *turn*, and *swipe* operations.
3. Open map and zoom $\times 2$ —In this task, the participants were asked to open a third-party map application and zoom $\times 2$ on the current location. The operations involved in this task included the *initial action*, *home*, *mental action*, *tap*, and *double tap* operations.
4. Turn on S Voice, then put to sleep—For this task, the participants were asked to initiate interaction with S Voice and then put the smartwatch to sleep. The operations involved in this task were the *initial action*, *home*, *mental action*, *double press*, and *gesture* operations.
5. Adjust Gear options—In this task, the participants were asked to access the Gear options and turn off the smartwatch. The following operations were utilised in this task: *initial action*, *home*, *mental action*, *press-and-hold*, and *tap*.

6.1.3. Procedure

The sessions were held in a quiet room and lasted 23.63 min on average. As in the previous studies, the participant was welcomed to the session and introduced to the study objective. This was followed by the presentation of an information sheet to answer anticipated questions and clarify the study procedure. Once consent was collected, the smartwatch was placed on the participant's non-dominant hand, where the participant was asked to feel free to adjust its position for comfort. The SJ4000 was then strapped to the participant's non-dominant shoulder and adjusted to guarantee a clear view of the smartwatch. This experiment used a mixed design approach, where each participant performed all tasks (within-subject) with only one group, either Apple Watch or Samsung Gear group (between-subject). The initial state for each participant was with the dominant and non-dominant arms to the sides when standing or on the lap when sitting down. This means that in the initial state, the smartwatch was in sleep mode. The order of the tasks was randomised for each participant. Each participant performed each given task until no errors were made. This is because the KLM and its variants were developed to model expert behaviour without any errors. Although the participants were instructed on how to perform the tasks, the hand posture when interacting with the smartwatch was not controlled. Each session was concluded with gratitude and compensation.

6.1.4. Data Analysis

All tasks were modelled with the proposed smartwatch model to predict their execution times. Each task was divided into its unit operators (see Table 6 for a sample model). The total time was then calculated, with the results shown in the same table. Model construction for each task began with the determination of the physical operators, without the mental action operators. These operators were later incorporated after the initial construction by following the modified heuristic rules for mental action operator insertion (see Section 4). The observed execution times were logged using BORIS [63] for each task. Each task began with the user in the initial state for the study, i.e., with the hands to the sides when standing or on the lap when sitting down. The end of the task was logged as soon as the final action or operation was completed. The average RMSPE between the observed and predicted execution times for each task was calculated to determine the prediction accuracy as follows:

$$PE_i = \frac{p_i - a_i}{a_i} \quad (1)$$

$$RMSPE = \sqrt{\text{mean}_{i=1,n}(100 \cdot |PE_i|)^2}$$

Here, a is the actual observed execution time, p is the predicted execution time, and n is the number of observations. This measure is often used in the literature to assess the validity of a revised KLM (e.g., [48,61,69]), following the analysis of the original KLM [27].

Table 6. Modelled sequence of operations for Apple Watch task 1 using the revised KLM for smartwatch interaction.

Operation	Operator	Unit Time (s)
Bring smartwatch, i.e., non-dominant hand, close for interaction	<i>Initial act</i>	0.82
Home dominant hand on smartwatch	<i>Home</i>	0.57
Mental preparation	<i>Mental act</i>	1.35
Force press on touchscreen to change watch face	<i>Tap and hold</i>	0.78
Swipe to second watch face option	<i>Swipe</i>	0.18
Swipe to third watch face options	<i>Swipe</i>	0.18
Mental preparation	<i>Mental act</i>	1.35
Tap on third option to change watch face	<i>Tap</i>	0.28
Execution time (s)		5.51

6.2. Findings

Table 7 and Figure 2 show the average observed times and the predicted execution times for each of the five tasks for the two smartwatches (Apple Watch Series 3 and Samsung Gear S3). Table 7 also shows the calculated RMSPE values for these tasks. The results show that the RMSPE values for all five tasks' execution times for each smartwatch are below the KLM's suggested RMSPE of 21%. However, it is also clear from Table 7 and Figure 2 that the revised model underestimates the predicted times for both the Apple Watch and Samsung Gear task groups. This underestimation can be explained by the artificial setting in which the task was completed. In most cases, participants were asked to repeat the task until confident they can be performed error free. This lead the participants to become cautious while completing the test tasks, and thus taking longer to complete. Moreover, these values are likely to represent upper-limit values for these tasks. This underestimation phenomenon was previously observed in other studies modelling mobile phones [57,61] and IVISs [47,50]. As detailed in the second study (see Section 5), the unit time for each operator was derived from users' interactions with an Apple Watch. Despite this, the revised model was able to accurately predict interaction times with a different type of smartwatch (Samsung Gear S3). This finding supports the generality of the model and its ability to model interactions with different types of smartwatches. The overall RMSPE was calculated to be 12.07%, thus validating the proposed model's adequacy in predicting execution times for smartwatch tasks.

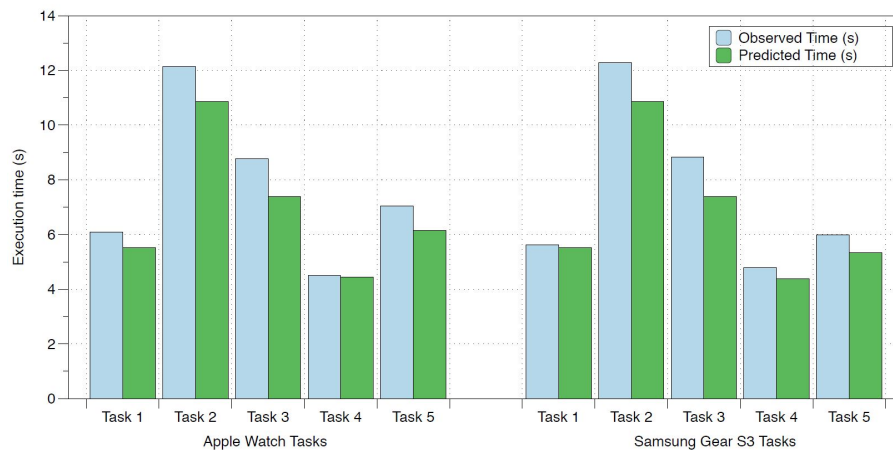


Figure 2. The average observed times and the predicted execution times for each of the five tasks for the two smartwatches (Apple Watch Series 3 and Samsung Gear S3).

Table 7. Average observed time, predicted execution time, and calculated RMSPE for each of the ten experimental task.

Smartwatch	Task#	Observed Time (s)	Predicted Time (s)	RMSPE (%)
Apple Watch	1	6.09	5.51	10.53
	2	12.14	10.87	10.22
	3	8.76	7.39	15.08
	4	4.51	4.44	11.80
	5	7.05	6.15	12.39
Apple Watch tasks' average RMSPE (%)				12.00
Samsung Gear S3	1	5.63	5.51	11.56
	2	12.28	10.87	11.01
	3	8.83	7.39	15.85
	4	4.79	4.39	10.93
	5	5.98	5.33	11.30
Samsung Gear tasks' average RMSPE (%)				12.13
Overall average RMSPE (%)				12.07

7. Conclusions and Future Work

In this paper, a revised KLM for smartwatch interaction (tentatively named Watch KLM) was developed to offer a pragmatic solution for designing and developing applications for smartwatches. Smartwatch interactions include touchscreen operations, button or key actions, and gestures. Model formulation and validation were achieved through three separate studies. The first study identified the range of actions that people perform when interacting with a smartwatch as a basis for revising the KLM relative to the original KLM [27] and the FLM [61]. The proposed model comprises fourteen operators that either were newly introduced or were adapted or retained from the original KLM [27] and FLM [61]. The model was kept as general as possible to ensure its ability to model most other types of smartwatches. To enable the practical use of the model, the unit time for each operator was computed in the second study (see Table 5). These two studies made it possible to construct models to predict execution times for smartwatch tasks. The final study was conducted to assess the revised model's predictive capabilities for realistically assessing smartwatch performance. The findings validate the model's accuracy with an average RMPSE of 12.07% across tasks; a value that falls within the acceptable range dictated by the original KLM.

Several limitations of these studies suggest possibilities for future work. Hand posture was observed in these studies, but its impact proved minimal and outside of this paper's scope. Further

work on hand placement on and around a smartwatch will be needed to refine the operators to more precisely consider individual behaviour. The operator unit times were derived in the second study from observations of participants interacting with an Apple Watch (see Section 5). To ensure precise predictions, plans are in place to expand this study with more participants and a wider variety of smartwatches. Typing actions were not considered in this paper and thus were excluded from computations. This, of course, leaves room for future research on various text input methods and their effects on the *tap* and *double tap* operations. The revised KLM for smartwatch interaction was validated in a controlled setting; thus, further studies will be needed to assess the model's performance in a natural setting. A final future direction of research is to assess the model's performance in real-life design cases by comparing the efficacy of different design choices to inform design decisions.

Funding: This research received no external funding.

Acknowledgments: I would like to thank all participants who took part in the studies.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

HCI	Human–computer interaction
GOMS	Goals, Operators, Methods, and Selection rules
KLM	Keystroke-level model
RMSPE	Root mean square percent error
FLM	Fingerstroke-level model
IVIS	In-vehicle information system
NFC	Near-field communication
TLM	Touch-level model

References

1. Starner, T. How wearables worked their way into the mainstream. *IEEE Pervasive Comput.* **2014**, *13*, 10–15. [CrossRef]
2. Kalantari, M. Consumers' adoption of wearable technologies: Literature review, synthesis, and future research agenda. *Int. J. Technol. Mark.* **2017**, *12*, 274–307. [CrossRef]
3. Hein, D.W.; Rauschnabel, P.A. Augmented reality smart glasses and knowledge management: A conceptual framework for enterprise social networks. In *Enterprise Social Networks*; Springer: Berlin, Germany, 2016; pp. 83–109.
4. Yoon, H.; Park, S.H.; Lee, K.T.; Park, J.W.; Dey, A.K.; Kim, S. A Case Study on Iteratively Assessing and Enhancing Wearable User Interface Prototypes. *Symmetry* **2017**, *9*, 114. [CrossRef]
5. Empatica Embrace. Available online: <https://www.empatica.com/en-eu/> (accessed on 29 April 2018).
6. Asimakopoulos, S.; Asimakopoulos, G.; Spillers, F. Motivation and user engagement in fitness tracking: Heuristics for mobile healthcare wearables. In *Informatics*; Multidisciplinary Digital Publishing Institute: Basel, Switzerland, 2017; Volume 4, p. 5.
7. Metcalf, D.; Milliard, S.T.; Gomez, M.; Schwartz, M. Wearables and the internet of things for health: Wearable, interconnected devices promise more efficient and comprehensive health care. *IEEE Pulse* **2016**, *7*, 35–39. [CrossRef] [PubMed]
8. Mardonova, M.; Choi, Y. Review of Wearable Device Technology and Its Applications to the Mining Industry. *Energies* **2018**, *11*, 547. [CrossRef]
9. Smartwatch Unit Sales Worldwide from 2014 to 2018. Available online: <https://www.statista.com/statistics/538237/global-smartwatch-unit-sales/> (accessed on 5 January 2018).
10. Rawassizadeh, R.; Price, B.A.; Petre, M. Wearables: Has the age of smartwatches finally arrived? *Commun. ACM* **2015**, *58*, 45–47. [CrossRef]
11. Cecchinato, M.E.; Cox, A.L. Smartwatches: Digital Handcuffs or Magic Bracelets? *Computer* **2017**, *50*, 106–109. [CrossRef]

12. Cecchinato, M.E.; Cox, A.L.; Bird, J. Always On (line)? User Experience of Smartwatches and their Role within Multi-Device Ecologies. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 3557–3568.
13. Guiard, Y. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *J. Motor Behav.* **1987**, *19*, 486–517. [[CrossRef](#)]
14. Schirra, S.; Bentley, F.R. It's kind of like an extra screen for my phone: Understanding Everyday Uses of Consumer Smart Watches. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; pp. 2151–2156.
15. Pizza, S.; Brown, B.; McMillan, D.; Lampinen, A. Smartwatch in vivo. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; pp. 5456–5469.
16. Liu, X.; Chen, T.; Qian, F.; Guo, Z.; Lin, F.X.; Wang, X.; Chen, K. Characterizing smartwatch usage in the wild. In Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, Niagara Falls, NY, USA, 19–23 June 2017; pp. 385–398.
17. Hara, K.; Umezawa, T.; Osawa, N. Effect of button size and location when pointing with index finger on smartwatch. In Proceedings of the International Conference on Human-Computer Interaction, Los Angeles, CA, USA, 2–7 August 2015; Springer: Berlin, Germany, 2015; pp. 165–174.
18. Lim, S.C.; Shin, J.; Kim, S.C.; Park, J. Expansion of Smartwatch Touch Interface from Touchscreen to Around Device Interface Using Infrared Line Image Sensors. *Sensors* **2015**, *15*, 16642–16653. [[CrossRef](#)] [[PubMed](#)]
19. Han, T.; Li, J.; Hasan, K.; Nakamura, K.; Gomez, R.; Balakrishnan, R.; Irani, P. PageFlip: Leveraging Page-Flipping Gestures for Efficient Command and Value Selection on Smartwatches. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; p. 529.
20. Xu, C.; Lyons, K. Shimmering smartwatches: Exploring the smartwatch design space. In Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction, Stanford, CA, USA, 16–19 January 2015; pp. 69–76.
21. Garcia, B.; Chu, S.L.; Nam, B.; Banigan, C. Wearables for Learning: Examining the Smartwatch as a Tool for Situated Science Reflection. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; p. 256.
22. Singh, G.; Delamare, W.; Irani, P. D-SWIME: A Design Space for Smartwatch Interaction Techniques Supporting Mobility and Encumbrance. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; p. 634.
23. Motti, V.G.; Caine, K. Micro interactions and Multi dimensional Graphical User Interfaces in the Design of Wrist Worn Wearables. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Los Angeles, CA, USA, 26–30 October 2015; SAGE Publications Sage CA: Los Angeles, CA, USA, 2015; Volume 59, pp. 1712–1716.
24. Chun, J.; Dey, A.; Lee, K.; Kim, S. A qualitative study of smartwatch usage and its usability. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2018**, *28*, 186–199. [[CrossRef](#)]
25. Choe, M.; Noh, G. Technology Acceptance of the Smartwatch: Health Consciousness, Self-Efficacy, Innovativeness. *Adv. Sci. Lett.* **2017**, *23*, 10152–10155. [[CrossRef](#)]
26. Hong, J.C.; Lin, P.H.; Hsieh, P.C. The effect of consumer innovativeness on perceived value and continuance intention to use smartwatch. *Comput. Hum. Behav.* **2017**, *67*, 264–272. [[CrossRef](#)]
27. Card, S.K.; Moran, T.P.; Newell, A. The Keystroke-level Model for User Performance Time with Interactive Systems. *Commun. ACM* **1980**, *23*, 396–410. [[CrossRef](#)]
28. Card, S.K.; Moran, T.P.; Newell, A. *The Psychology of Human-Computer Interaction*; L. Erlbaum Associates Inc.: Hillsdale, NJ, USA, 1983.
29. Holleis, P.; Otto, F.; Hussmann, H.; Schmidt, A. Keystroke-level model for advanced mobile phone interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI '07, San Jose, CA, USA, 30 April–3 May 2007; pp. 1505–1514.
30. El Batran, K.; Dunlop, M.D. 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), Toronto, ON, Canada, 23–26 September 2014; pp. 283–286.

31. Pettitt, M.; Burnett, G.; Stevens, A. An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI '07, San Jose, CA, USA, 30 April–3 May 2007; pp. 1515–1524.
32. Elwart, T.; Green, P.; Lin, B. *Predicting Driver Distraction Using Computed Occlusion Task Times: Estimation of Task Element Times and Distributions*; Technical Report; The University of Michigan Transportation Research Institute: Ann Arbor, MI, USA; 2015.
33. Moreta, E. A Predictive Model for User Performance Time with Natural User Interfaces Based on Touchless Hand Gestures. Ph.D. Thesis, Universidad de Chile, Santiago, Chile, 2016.
34. Bevan, N.; Carter, J.; Harker, S. ISO 9241-11 revised: What have we learnt about usability since 1998? In *International Conference on Human-Computer Interaction*; Springer: Berlin, Germany, 2015; pp. 143–151.
35. Mo, F.; Yi, S.; Zhou, J. Effect of Icon Amount and Visual Density on Usability of Smartwatches. In *International Conference on Human Aspects of IT for the Aged Population*; Springer: Berlin, Germany, 2016; pp. 466–477.
36. Wong, S.; Yang, L.; Riecke, B.; Cramer, E.; Neustaedter, C. Assessing the usability of smartwatches for academic cheating during exams. In Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services, Vienna, Austria, 4–7 September 2017; p. 31.
37. Roberts, T.L.; Moran, T.P. The evaluation of text editors: Methodology and empirical results. *Commun. ACM* **1983**, *26*, 265–283. [[CrossRef](#)]
38. Olson, J.R.; Nilsen, E. Analysis of the Cognition Involved in Spreadsheet Software Interaction. *Hum. Comput. Interact.* **1987**, *3*, 309–349. [[CrossRef](#)]
39. Edwards, A.D.N. Modelling blind users' interactions with an auditory computer interface. *Int. J. Man-Mach. Stud.* **1989**, *30*, 575–589. [[CrossRef](#)]
40. Lane, D.; Napier, H.A.; Batsell, R.; Naman, J. Predicting the Skilled Use of Hierarchical Menus With the Keystroke-Level Model. *Hum. Comput. Interact.* **1993**, *8*, 185–192. [[CrossRef](#)]
41. Quintana, Y.; Kamel, M.; McGeachy, R. Formal Methods for Evaluating Information Retrieval in Hypertext Systems. In Proceedings of the 11th Annual International Conference on Systems Documentation, Waterloo, ON, Canada, 5–8 October 1993; pp. 259–272. [[CrossRef](#)]
42. Schrepp, M. GOMS Analysis As a Tool to Investigate the Usability of Web Units for Disabled Users. *Univ. Access Inf. Soc.* **2010**, *9*, 77–86. [[CrossRef](#)]
43. Lee, A. Exploring User Effort Involved in Using History Tools Through MHP/GOMS: Results and Experiences. *Hum. Comput. Interact.* **1995**, 109–114. [[CrossRef](#)]
44. Burns, M.T.; Ritter, F.E. Using Naturalistic Typing to Update Architecture Typing Constants. In Proceedings of the International Conference on Cognitive Modeling, University Park, PA, USA, 3–5 August 2016; pp. 169–174.
45. Manes, D. *Evaluation of a Driver Interface: Effects of Control Type (Knob versus Buttons) and Menu Structure (Depth versus Breadth)*; Technical Report; The University of Michigan Transportation Research Institute: Ann Arbor, MI, USA, 1997.
46. Green, P. Estimating Compliance with the 15-Second Rule for Driver-Interface Usability and Safety. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Los Angeles, CA, USA, 1 September 1999; pp. 987–991. [[CrossRef](#)]
47. Pettitt, M.; Burnett, G. Visual Demand Evaluation Methods for In-Vehicle Interfaces. *Int. J. Mob. Hum. Comput. Interact.* **2010**, *2*, 45–57. [[CrossRef](#)]
48. Schneegaß, S.; Pfleging, B.; Kern, D.; Schmidt, A. 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), Pittsburgh, PA, USA, 11–12 November 2010; ACM: New York, NY, USA, 2011; pp. 71–78.
49. Manes, D. *Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models*; Technical Report; The University of Michigan Transportation Research Institute: Ann Arbor, MI, USA, 1997.
50. Green, P.; Kang, T.P.; Lin, B. *Touch Screen Task Element Times for Improving SAE Recommended Practice J2365: First Proposal*; Technical Report; The University of Michigan Transportation Research Institute: Ann Arbor, MI, USA, 2015.
51. Dunlop, M.D.; Crossan, A. Predictive text entry methods for mobile phones. *Pers. Technol.* **2000**, *4*, 134–143. [[CrossRef](#)]

52. Liu, Y.; R  ih  , K.J. Predicting Chinese Text Entry Speeds on Mobile Phones. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI '10, Atlanta, GA, USA, 10–15 April 2010; ACM: New York, NY, USA, 2010; pp. 2183–2192. [\[CrossRef\]](#)
53. Li, F.C.Y.; Guy, R.T.; Yatani, K.; Truong, K.N. The 1line keyboard: A QWERTY layout in a single line. In Proceedings of the 24th annual ACM symposium on User interface software and technology, Santa Barbara, CA, USA, 16–19 October 2011; pp. 461–470.
54. Myung, R. Keystroke-level analysis of Korean text entry methods on mobile phones. *Int. J. Hum. Comput. Stud.* **2004**, *60*, 545–563. [\[CrossRef\]](#)
55. Cao, S.; Ho, A.; He, J. Modeling and predicting mobile phone touchscreen transcription typing using an integrated cognitive architecture. *Int. J. Hum. Comput. Interact.* **2018**, *34*, 544–556. [\[CrossRef\]](#)
56. Cox, A.L.; Cairns, P.A.; Walton, A.; Lee, S. Tlk or txt? Using voice input for SMS composition. *Pers. Ubiquitous Comput.* **2008**, *12*, 567–588. [\[CrossRef\]](#)
57. Holleis, P.; Scherr, M.; Broll, G. A Revised Mobile KLM for Interaction with Multiple NFC-tags. In Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction—Volume Part IV (INTERACT'11), Lisbon, Portugal, 5–9 September 2011; Springer-Verlag: Berlin/Heidelberg, Germany, 2011; pp. 204–221.
58. Li, H.; Liu, Y.; Liu, J.; Wang, X.; Li, Y.; Rau, P.L.P. Extended KLM for Mobile Phone Interaction: A User Study Result. In Proceedings of the CHI '10 Extended Abstracts on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; pp. 3517–3522. [\[CrossRef\]](#)
59. Rice, A.D.; Lartigue, J.W. Touch-level Model (TLM): Evolving KLM-GOMS for Touchscreen and Mobile Devices. In Proceedings of the 2014 ACM Southeast Regional Conference, Kennesaw, GA, USA, 28–29 March 2014; p. 53. [\[CrossRef\]](#)
60. El Batran, K.; Dunlop, M.D. 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, MobileHCI '14, Toronto, ON, Canada, 23–26 September 2014; ACM: New York, NY, USA, 2014; pp. 283–286. [\[CrossRef\]](#)
61. Lee, A.; Song, K.; Ryu, H.B.; Kim, J.; Kwon, G. Fingerstroke time estimates for touchscreen-based mobile gaming interaction. *Hum. Mov. Sci.* **2015**, *44*, 211–224. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Al-Megren, S.; Altamimi, W.; Al-Khalifa, H.S. Blind FLM: An Enhanced Keystroke-Level Model for Visually Impaired Smartphone Interaction. In *IFIP Conference on Human-Computer Interaction*; Springer: Berlin, Germany, 2017; pp. 155–172.
63. Behavioral Observation Research Interactive Software (BORIS). Available online: <http://www.boris.unito.it/> (accessed on 17 September 2016).
64. Ueno, K.; Go, K.; Kinoshita, Y. Investigation of smartwatch touch behavior with different postures. In Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia, Stuttgart, Germany, 26–29 November 2017; pp. 57–61.
65. Kieras, D. *Using the Keystroke-Level Model to Estimate Execution Times*; University of Michigan: Ann Arbor, MI, USA, 2001; Volume 555.
66. Human Interface Guidelines—Buttons. Available online: <https://developer.apple.com/watchos/human-interface-guidelines/interface-elements/buttons/> (accessed on 30 April 2017).
67. Esteves, M.; Komischke, T.; Zapf, S.; Weiss, A. Applied user performance modeling in industry—A case study from medical imaging. In *International Conference on Digital Human Modeling*; Springer: Berlin, Germany, 2007; pp. 576–585.
68. Erazo, O.; Pino, J.A. Predicting user performance time for hand gesture interfaces. *Int. J. Ind. Ergon.* **2018**, *65*, 122–138. [\[CrossRef\]](#)
69. Erazo, O.; Pino, J.A. Predicting task execution time on natural user interfaces based on touchless hand gestures. In Proceedings of the 20th International Conference on Intelligent User Interfaces, Atlanta, GA, USA, 29 March–1 April 2015; pp. 97–109.

