Gesture-Controlled User Input to Complete Questionnaires on Wrist-Worn Watches

Oliver Amft[1,2], Roman Amstutz[1,3], Asim Smailagic[3], Dan Siewiorek[3], and Gerhard Tröster[1]

[1]Wearable Computing Lab., ETH Zurich, CH-8092 Zurich, Switzerland
[2]Signal Processing Systems, TU Einhoven, 5600 MB Einhoven, The Netherlands
[3]Carnegie Mellon University, Pittsburgh, PA 15213, USA
{amft,troester}@ife.ee.ethz.ch, {asim,dps}@cs.cmu.edu

Abstract. The aim of this work was to investigate arm gestures as an alternative input modality for wrist-worn watches. In particular we implemented a gesture recognition system and questionnaire interface into a watch prototype. We analyzed the wearer's effort and learning performance to use the gesture interface and compared their performance to a classical button-based solution. Moreover we evaluated the system performance to spot wearer gestures and the system responsiveness. Our wearer study showed that the watch achieved a recognition accuracy of more than 90%. Completion times showed a clear decrease from 3 min in the first repetition to 1 min, 49 sec in the last one. Similarly, variance of completion times between wearers decreased during repetitions. Completion time using the button interface was 36 sec. Ratings of physical and concentration effort decreased during the study. Our results confirm that wearer training state is rather reflected in completion time than recognition performance.

Keywords: gesture spotting, activity recognition, eWatch, user evaluation.

1 Introduction

Gesture-based interfaces have been proposed as alternate modality for controlling stationary computers, e.g. to navigate in office applications and play immersive console games. In contrast to their classic interpretation to support conversation, gestures are understood in this area as directed body movements, primarily of arms and hands, to interact with computers. Gesture-based interfaces can enrich and diversify interaction options as in gaming. Moreover, they are vital for computer access by handicapped, such as enabled by sign language interfaces, and for further applications and environments where traditional computer interaction methods are not acceptable.

Mobile systems and devices are a primary field concern for such alternate interaction solutions. While the above cited stationary applications demonstrate the applicability of directed gestures for interaction, future mobile solutions could

profit from deployment of gesture-based interfaces in particular. Currently, mobile systems lack solutions that minimize user attention or support access for users with specific interaction needs. Hence, gestures that are directly sensed and recognized by a mobile or wearable device are of common interest for numerous applications.

In this work, we investigate gesture-based interaction using a wrist-worn watch device. We consider gestures as an intuitive modality, especially for watches, and potentially feasible for wearers that cannot operate tiny watch buttons. To this end, it is essential to evaluate the wearer's performance and convenience while operating such an interface. Moreover, the integration of gesture interfaces into a watch device has not been previously evaluated. Resource constraints of wrist-worn watch devices impose challenging restrictions regarding processing complexity for embedding a gesture recognition solution into watch devices.

Consequently, this paper provides the following contributions:

1. We present a prototype of an intelligent wrist-worn watch, the eWatch, and demonstrate that a recognition procedure particularly designed for gesture spotting, can be embedded into this device. The recognition procedure consists of two stages to spot potential gestures in continuous acceleration data, and classify the type of gesture. Feasibility of this recognition procedure was assessed by an analysis of the implementation requirements.

2. We present a user study evaluating the wearer's performance in executing gestures to complete a questionnaire that was implemented on the watch as well. In particular, we investigated recognition accuracy and wearer learning effects during several repetitions of completing the questionnaire. Moreover, we compared the time required to complete the questionnaire using the gesture interface to a button interface.

As this work evaluates gesture-based interaction regarding both, technical feasibility and user performance, it provides an novel insight into the advantages and limitations of gesture interfaces. We believe that these results are generally relevant for gesture-operated mobile systems.

Section 2 discusses related works and approaches to develop intelligent watches, gesture-operated mobile devices, and recognition procedures for gesture spotting. Subsequently, Sections 3 and 4 briefly present the watch system and the embedded gesture recognition procedure, respectively. The user study and evaluation results are presented in Section 5. Section 6 concludes on the results of this work.

2   Related Work

Wrist-worn watches have been proposed as truly wearable processing units. A pioneering development was the IBM Linux Watch [1]. Besides time measurement, various additional applications of wristwatches have been identified and brought to commercial success. This includes sports and fitness monitoring and support watches as, e.g. realized by Polar (www.polar.fi/en/) and Suunto (www.suunto.com). With the Smart Personal Object Technology(SPOT)[2],

consumer watches become broadcast news receivers. Similarly, wristwatches have been used as a mobile phone (www.vanderled.com) or for GPS-navigation (www.mainnav.com). Besides the frequent button-based control, wristwatches have been equipped with touch-sensitive displays (www.tissot.ch) to improve interaction. No related work was identified that investigated gesture-based interaction for wristwatches as it is proposed in this paper.

Gesture recognition has been investigated for various applications in areas, such as activity recognition and behavior inference [3,4,5], immersive gaming [6,7], and many forms of computer interaction. In this last category, systems have been proposed to replace classical computer input modalities. A review on the various applications was compiled by Mitra and Acharya [8]. In this work, we focus our discussion on related approaches in gesture-operated mobile devices. Moreover, we provide a coarse overview on established gesture recognition and spotting techniques.

2.1   Gesture-Operated Mobile Devices

Gesture spotting and recognition based on body-worn sensors has primarily used accelerometers to identify body movement patterns. These sensors are found in many current mobile phones. However, due to the constraint processing environment of watches, their interfaces had classically been restricted to simple button-based solutions. Consequently, gesture interfaces for watches have not been extensively investigated.

Recent investigations started to address the implementation challenge of gesture interfaces onto mobile devices, beyond simple device turning moves. Kallio et al. [9] presented an application using acceleration sensors embedded in a remote control to manage home appliances. Their work was focused on confirming the feasibility for classifying different gestures using hidden Markov models (HMMs). Recently, Kratz and Ballagas [10] presented a pervasive game that relied on gestures as input recognized on mobile phones.

2.2   Gesture Spotting

Various algorithms have been proposed for spotting and classifying gestures. While the first task relates to the identification of gestures in a continuous stream of sensor data, the second task deals with the discrimination of particular gesture types. The recognition procedure must be capable of excluding non-relevant gestures and movements. For the spotting task various methods have been presented that cope with the identification problem, e.g. [11,12,5]. We deploy in this work an approach related to the work of Lee and Kim [11]. The authors have used the Viterbi algorithm to preselect relevant gestures. For the classification task, many works have proposed HMMs, e.g. [11,5]. For the implementation presented this work we followed this approach by deriving individual HMMs for each gesture class and used a threshold model to discriminate non-relevant movements.

3   Watch System and Embedded Questionnaire

We used in this investigation an intelligent watch prototype, the eWatch. Figure 1 shows the device running a questionnaire application. The eWatch consists of a ARM7 processor without float-point unit, running with up to 80 MHz. A detailed description of the system architecture can be found in [13]. In this work, we used the MEMS 3-axes accelerometer that is embedded in the eWatch, to sense acceleration of the wearer's arm and supply the recognition procedure with sensor data.

Fig. 1. eWatch prototype running a questionnaire application

The questionnaire was chosen as an evaluation and test application to verify that the gesture recognition procedure achieves an user-acceptable recognition rate. Moreover, we used the questionnaire to stimulate the wearer to perform alternating gestures during the interface evaluation in Section 5.

The questionnaire application was designed to display a question on the left side of the watch screen and provides four answer options on the right side to choose from. In order to respond to the question, the wearer had to perform at least one "select" gesture for each question. This gesture would choose the highlighted answer and advance to the next question dialog. When the wearer intended to choose a different answer than the currently selected one, "scroll-up" and "scroll-down" gestures could be used to navigate between possible answers. Figure 2 shows the individual gestures considered in this evaluation. The "scroll-up" and "scroll-down" gestures can be described as outward and inward (towards the trunk) rotation movements of the arm. The "select" gesture consisted of raising and lowering the arm two times.

Fig. 2. Gestures used to operate the eWatch device: "scroll-up", "scroll-down", and "select"

These gestures were selected empirically out of 13 different gestures repeatedly performed by nine test persons. The gestures were chosen based on initial tests of the recognition procedure, as detailed in Section 4 below, and according to qualitative feedback of the test persons. The reliable spotting and classification were however given priority, since we considered an accurate operation as most essential design goal.

Although related gesture recognition evaluations considered far larger gesture sets successfully, e.g. in the work of Lee and Kim [11], we expected that additional gesture options would be confusing for the wearer. In addition, a larger set of gestures may require longer training times for the user.

4   Watch-Based Gesture Spotting

In order to evaluate gesture-based interaction for a wristwatch, we developed and implemented a recognition procedure into the eWatch device. We briefly summarize the design and implementation results here, which indicate the feasibility of the gesture recognition approach.

4.1   Gesture Recognition Procedure

The recognition procedure consists of two distinct stages: the spotting of relevant gestures that are used to operate the questionnaire, and the classification of these gestures. The first stage has to efficiently process the continuous stream of sensor data and identify the gestures embedded in arbitrary other movements. Due to this search, this task can have a major influence on the processing requirements. The second stage evaluates the selected gestures and categorizes them according to individual pattern models. The deployed procedure is briefly summarized below.

For the spotting task in this work, we extracted the dominating acceleration axis, defined as the axis with the largest amplitude variation within the last five sampling points. The derivative of this acceleration was used in combination with a fixed sliding window to spot gestures. Individual discrete left-right HMMs for each gesture were manually constructed, with six states for the scroll gestures, nine states for the select gesture. A code-book of 13 symbols was used to represent the derivative amplitude in strong/low increase/decrease for all acceleration axes and rest, for small amplitudes. The Viterbi algorithm was used to calculate the most probable state transition sequence for the current sliding window. The end of a gesture was detected if the end state in an HMM was reached in the current window. Using this end-point, the corresponding gesture begin was determined. Based on these preselected gestures, a HMM-based classification was applied. A gesture was classified to the HMM achieving the maximum likelihood value. Moreover, the gesture was retained only if the likelihood value exceeded that of a threshold HMM. The threshold HMM was derived according to [11]. If more than one gesture was detected in one sliding window, the one with larger likelihood was retained.

4.2   Watch Implementation

Our implementation of the gesture recognition procedure used the eWatch base system for managing hardware components. A CPU clock of 65 MHz was used. Acceleration data was sampled at 20 Hz. The HMMs required a total memory of approximately 3.75 KB.

Our analysis showed a processing time below 1 ms for each HMM at a sliding window size of 30 samples. The total gesture recognition delay was below 2.7 ms for the entire procedure. This delay remains far below delays that could be noticed by the user. These results confirm the applicability of the recognition method.

Our empirical analysis of the gesture recognition performance showed that the "scroll-down" gesture perturbed the overall recognition performance. By restricting the interface to "scroll-up" and "select" gestures, we improved robustness, while simplifying the interface. The questionnaire application was equipped with a wrap-around feature, in order to select each answer option by just using "scroll-up". Our subsequent user evaluation, as detailed below, confirmed that this choice did not restrict the users in operating the application.

5   User Evaluation of Gesture and Button Interfaces

We conducted a user study to evaluate the feasibility of the gesture-operated interface and to assess the user's performance, perception, and comfort using a gesture interface. In particular, we investigated recognition accuracy and wearer learning effects during several repetitions of completing the questionnaire. Finally, we compared the time required to complete the questionnaire using the gesture interface to a classic button interface.

5.1   Study Methodology

Ten students were recruited to wear the eWatch and complete the questionnaire in four repetitions. Five users from non-technical background were included in order to analyze whether expertise with technical systems would influence performance or ratings.

The users performed the evaluation individually. Initially, they watched a training video demonstrating the handling of the watch and the gestures. If they had further questions, these were resolved afterwards. Subsequently, the users attached the watch and performed four repetitions of a questionnaire asking for eight responses. This questionnaire was designed to indicate a particular answer option for each question, which the users were asked to select. With this protocol, the number of gestures performed by each user was maintained comparable.

For each repetition of the questionnaire the completion time was measured. Each gesture execution, conducted gesture, and result was logged by an observer. In addition, the recording sessions were videotaped using a video camera installed in the experiment room. The camera was positioned over the head of the users at the side of the arm, where the watch was worn. In this way, the video captured

scene, gestures performed by the user, as well as the response shown at the watch screen when users turned the watch to observe the result. This video was later used as a backup for the experiment observer to count correct gesture performances.

After each repetition users completed an intermediate paper-based assessment questionnaire, assessing their qualitative judgment of convenience to use the system, rating their personal performance, physical effort, concentration effort, and performance of the system. After all repetitions were completed, the users completed additional assessment questions. These questions were intended to capture general impressions of the gesture interface.

Physical effort was assessed by asking the question: "How tired are you from performing the gestures?". A visual analog scale (VAS) was used with 1 (not tired at all) to 10 (very tired). Concentration effort was assessed with the question: "How much did you have to think about how the gesture has to be performed?". A VAS with 1 (not at all) and 10 (very much) was used.
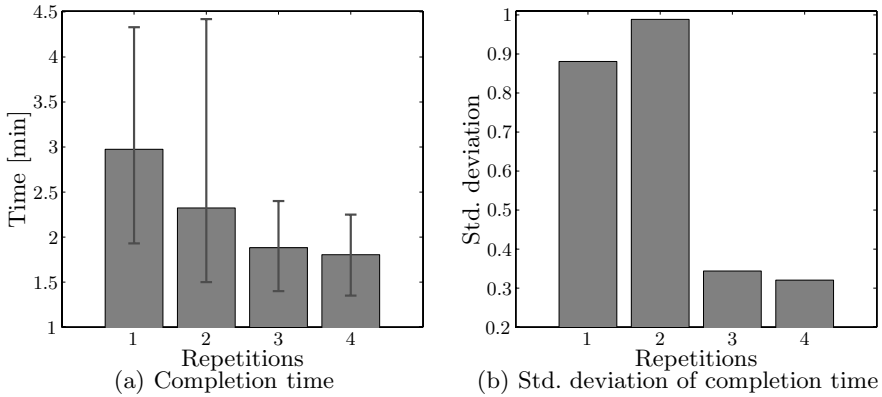
After all repetitions using the gesture interface, the users were asked to complete the same watch questionnaire application once more using an eWatch with a button interface. Finally, the users were asked to rate the gesture and button-based interfaces based on their experience made in the evaluation.

5.2   User Study Results

The recognition accuracy was evaluated by comparing the user-performed gestures and reactions to the eWatch screen feedback. Figure 3 shows the average accuracies for each questionnaire repetition and both gestures: "scroll-up" and "select". Accuracy was derived here as the ratio between recognized and total performed gestures. From these average accuracy results, no clear learning effect


(a) Gesture "scroll-up"   (b) Gesture "select"
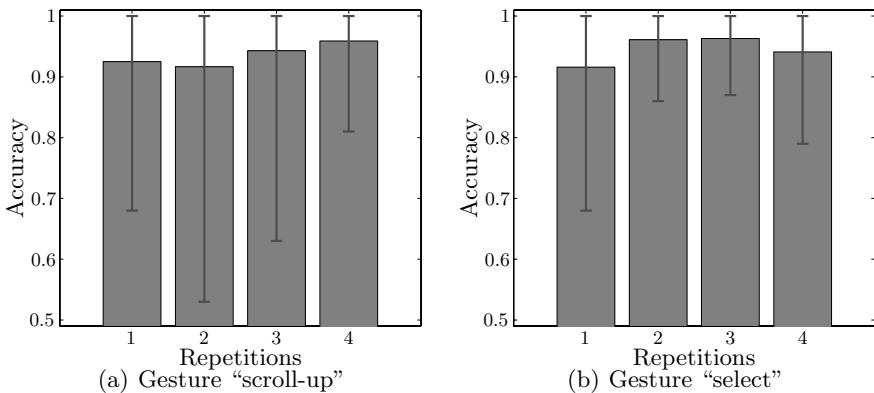
Fig. 3. Average recognition accuracies for all four repetitions of the study using the gesture interface. Error bars indicate minimum and maximum recognition performances individual users.

(a) Completion time   (b) Std. deviation of completion time

Fig. 4. Average user completion times for all four repetitions of the study using the gesture interface. Error bars indicate minimum and maximum times for individual users.

can be observed. Large initial accuracies above 90% indicate that the initial the user training, before starting the evaluation allowed the users to acquire a good skill in performing the gestures. Overall, the accuracies remained above 90% for all repetitions, while individual performances for "scroll-up" increased during the last three repetitions. The drop in the performance for "select" might have been caused by the fact that this gesture required more effort to be performed. Consequently, users may have become tired.

A clear trend can be observed from the questionnaire completion times shown in Figure 4. Both absolute time required, as well as the std. deviation for all

(a) Physical effort rating   (b) Concentration rating
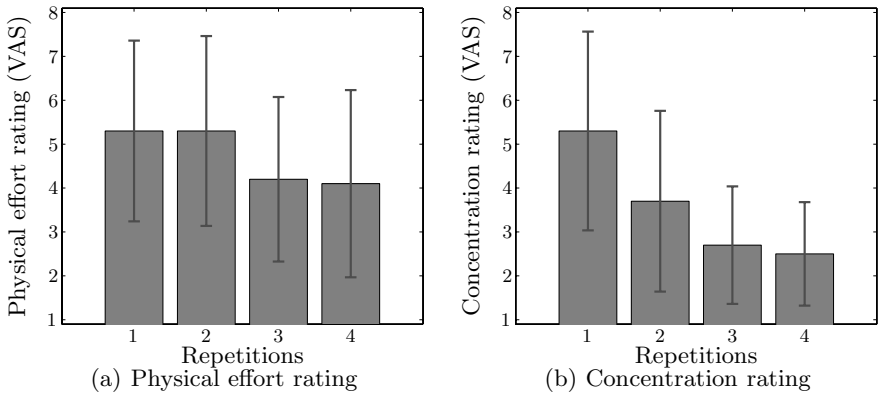
Fig. 5. User ratings of physical effort and concentration for all four repetitions of the study using the gesture interface. Error bars indicate std. deviations for individual users. The ratings were obtained using a VAS from 1 (low effort) to 10 (high effort). See Section 5 for a detailed description.

users decreased during the study, reaching a minimum in the last repetition. This result shows that completion time can indicate user accommodation to the gesture interface and improved training state.

Furthermore, we assessed the completion time for the button interface. Average completion time was here $\sim$36 sec, with a std. deviation of $\sim$2.5 sec. In comparison to the last repetition using the gesture interface, which was performed immediately beforehand, a three times lower completion time can be observed. As the gesture recognition and watch reaction time was confirmed to be not noticeable to the user, this difference can be entirely attributed to time required to perform the gestures.

Figure 5 shows the results from the user ratings on VAS between 1 (low effort) to 10 (high effort). The average ratings of physical effort and concentration decreased over all repetitions. Only three users reported an constant or increasing physical effort to perform the gestures. These results support our assumption of an improvement in the user training state during the repetitions.

6   Discussion and Conclusions

Our investigation confirmed that a gesture interface deployed into a watch device is a feasible approach to replace classic button-style interaction. The evaluations performed in this work indicate that a gesture interface requires training, even for a very limited number of gestures. Consequently even after four repetitions of the questionnaire application considered in this work, we observed an improvement in the user performance.

This user training state was not clearly reflected in an improvement of gesture recognition accuracy. However, the completion time needed to achieve the task was found in this study to be directly related to the training. Hence, with improved training state users required less time to perform the gestures. This improved training state was confirmed by user ratings of required physical effort and concentration on the task. Both metrics decreased during repetitions of the questionnaire application.

Our gesture-based interaction concept did not meet the low completion times of a comparable button-based solution. This observation was confirmed by final user ratings for a personally preferred interface. Nine out of ten users preferred the button-based interface. While our study was successful to evaluate the gesture interface itself, we expect that this disadvantageous rating for the gesture-based interaction can be explained by the questionnaire task and setting chosen for this investigation. Users were asked to perform the task, without further constraints in an isolated lab environment. Although this was a useful methodology for this evaluation stage, we expect that gesture-based interaction, can be a very vital alternative in particular applications and contexts. Several potential applications exist, in which buttons cannot be used, including interaction for handicapped individuals that cannot use small buttons as well as interaction in work environments, where the worker cannot use hands or wears gloves. These vital application areas and user groups should be considered further, based on the successful results obtained in this work.

References

1. Narayanaswami, C., Raghunath, M., Kamijoh, N., Inoue, T.: What would you do with a hundred mips on your wrist? Technical Report RC 22057 (98634), IBM Research (January 2001)
2. Goldstein, H.: A dog named spot (smart personal objects technology watch). IEEE Spectrum 41(1), 72–73 (2004)
3. Ivanov, Y., Bobick, A.: Recognition of visual activities and interactions by stochastic parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(8), 852–872 (2000)
4. Kahol, K., Tripathi, K., Panchanathan, S.: Documenting motion sequences with a personalized annotation system. IEEE Multimedia 13(1), 37–45 (2006)
5. Junker, H., Amft, O., Lukowicz, P., Tröster, G.: Gesture spotting with body-worn inertial sensors to detect user activities. Pattern Recognition 41(6), 2010–2024 (2008)
6. Bannach, D., Amft, O., Kunze, K.S., Heinz, E.A., Tröster, G., Lukowicz, P.: Waving real hand gestures recorded by wearable motion sensors to a virtual car and driver in a mixed-reality parking game. In: Blair, A., Cho, S.B., Lucas, S.M. (eds.) CIG 2007: Proceedings of the 2nd IEEE Symposium on Computational Intelligence and Games, April 2007, pp. 32–39. IEEE Press, Los Alamitos (2007)
7. Schlömer, T., Poppinga, B., Henze, N., Boll, S.: Gesture recognition with a wii controller. In: TEI 2008: Proceedings of the 2nd international conference on Tangible and embedded interaction, pp. 11–14. ACM, New York (2008)
8. Mitra, S., Acharya, T.: Gesture recognition: A survey. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 37(3), 311–324 (2007)
9. Kallio, S., Kela, J., Korpipää, P., Mäntyjärvi, J.: User independent gesture interaction for small handheld devices. International Journal of Pattern Recognition and Artificial Intelligence 20(4), 505–524 (2006)
10. Kratz, S., Ballagas, R.: Gesture recognition using motion estimation on mobile phones. In: PERMID 2007: Proceedings of the 3rd International Workshop on Pervasive Mobile Interaction Devices, Workshop at the Pervasive 2007 (May 2007)
11. Lee, H.K., Kim, J.H.: Gesture spotting from continuous hand motion. Pattern Recognition Letters 19(5-6), 513–520 (1998)
12. Deng, J., Tsui, H.: An HMM-based approach for gesture segmentation and recognition. In: ICPR 2000: Proceedings of the 15th International Conference on Pattern Recognition, September 2000, vol. 2, pp. 679–682 (2000)
13. Maurer, U., Rowe, A., Smailagic, A., Siewiorek, D.: eWatch: A wearable sensor and notfication platform. In: BSN 2006: Proceedings of the IEEE International Workshop on Wearable and Implantable Body Sensor Networks, Washington, DC, USA, pp. 142–145. IEEE Computer Society Press, Los Alamitos (2006)