# Investigation of User Rating Behavior Depending on Interaction Methods on Smartphones


Shabnam Najafian
TU München
Boltzmannstr. 3
85748 Garching
Germany
s.najafian@tum.de

Wolfgang Wörndl
TU München
Boltzmannstr. 3
85748 Garching
Germany
woerndl@in.tum.de

Beatrice Lamche
TU München
Boltzmannstr. 3
85748 Garching
Germany
lamche@in.tum.de



## ABSTRACT

Recommender systems are commonly based on user ratings to generate tailored suggestions to users. Instabilities and inconsistencies in these ratings cause noise, reduce the quality of recommendations and decrease the users' trust in the system. Detecting and addressing these instabilities in ratings is therefore very important. In this work, we investigate the influence of interaction methods on the users' rating behavior as one possible source of noise in ratings. The scenario is a movie recommender for smartphones. We considered three different input methods and also took possible distractions in the mobile scenario into account. In a conducted user study, participants rated movies using these different interaction methods while either sitting or walking. Results show that the interaction method influences the users' ratings. Thus, these effects contribute to rating noise and ultimately affect recommendation results.



## Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces - Input devices and strategies, Interaction styles

## General Terms

Design, Experimentation, Human Factors.

## Keywords

user interfaces, recommender systems, rating behavior, user study, gestural interaction, mobile applications.


## 1. INTRODUCTION

In an age where information overload is becoming greater, generating accurate recommendations plays an increasingly important role in our everyday life. On the other hand, smartphones equipped with some set of embedded sensors provide an important platform to access data. Moreover, limitations in the user interface and the absence of suitable interaction methods makes it more and more difficult for mobile users to filter necessary information. Personalization and customization of the generated data helps deal with this information overload. Recommendation techniques are a subarea of intelligent personalizing and are seeking to obtain the users' preferences to allow personalized recommendations of tailored items. Recommender systems apply various recommendation techniques such as collaborative filtering, content-based, hybrid or context-aware recommendations, but all depend on acquiring accurate preferences (e.g. ratings) from users.

Preference acquisition is addressed via either explicit (user states his/her preferences), or implicit (system observes and analyzes the user's behavior) methods [3]. Because of the ambiguous nature of the implicit approach, explicit techniques are often employed to gather more reliable ratings from users to capture the users' preferences. Existing research usually assume stable ratings, i.e. the assumptions is that an available rating exactly reflect the user's opinion about an item. However, explicitely entered ratings may contain some level of noise. If this is the case, the system can not generate accurate recommendations. A lot of reasearch has been invested to increase the accuracy of recommendation algorithms, but relatively little to investigate the rating process.

This work explores one probable source of error in the rating process on smartphones which has not been considered much yet: the influence of input methods on the resulting ratings. Our specific scenario is a recommender system on a mobile device (smartphone). Mobile applications offer different input options for interaction including touchscreen and free-form gestures [7]. Touchscreen gestures allow users to tap on the screen, either using on-screen buttons or other interface elements, e.g. sliders. Free-form gestures do not require the user to actively touch the screen but to move the devices to initiate functions. In our previous work, we investigated which interaction methods are preferrable from a user's perspective for certain recommender system tasks [8].

The aim of this user study was to show that participants rate items differently depending on the applied input method. Errors that may occur due to re-rating were also taken into account to reduce other noises. We considered two situations in our study: the user were either sitting and concentrated on the task, or walking around and thus possibly distracted by the environment. We also measured the


*IntRS 2014*, October 6, 2014, Silicon Valley, CA, USA.

ease of use and effectiveness of our implementation based on an online survey.

The rest of the paper is organized as follows. We first outline related work. Next, we present our employed interaction methods and their implementation. In Section 4, we explain the setup and the results of our user study. Finally, we conclude the paper with a summary and a brief outlook.

## 2. RELATED WORK

Analyzing and characterizing noise in user rating of recommender systems in order to improve the quality of recommendations and therefore user acceptance is still an open research problem. Jawaheer et al. [3] recently surveyed methods to model and acquire user prefereces for recommender systems, distinguishing between explicit and implicit methods. They also mention that user ratings inherently have noise and cited some earlier studies. One earlier example is the study by Cosley et al. [2]. They investigated the influence of showing rating predictions when asking users to re-rate items. They found out that users applied their original rating more often when shown the predictions.

Amatriain et al. [1] attempted to quantify the noise due to inconsistencies of users in giving their feedback. They examined 100 movies from the Netflix Prize database in 3 trials of the same task: rating 100 movies via a web interface at different points in time. RMSE values were measured in the range of 0.557 and 0.8156 and four factors influencing user inconsistencies: 1) Extreme rating are more consistent were inferred, 2) Users are more consistent when movies with similar ratings are grouped together, 3) The learning effect on the setting improves the user's assessment, 4) The faster act of clicking on user's part does not yield more inconsistencies.

Nguyen et al. [5] performed a re-rate experiment consisting of 386 users and 38586 ratings in MovieLens. They developed four interfaces: one with minimalistic support that serves as the baseline, one that shows tags, one that provides exemplars, and another that combines the previous two features, to address two possible source of errors within the rating method. The first assumption is that users may not clearly recall items. Secondly, users may struggle to consistently map their internal preferences to the rating scale. The results showed that although providing rating support helps users rate more consistently, participants liked baseline interfaces because they perceived the interfaces to be more easy to use. Nevertheless, among interfaces providing rating support, the proposed one that provides exemplars appears to have the lowest RMSE, the lowest minimum RMSE, and the least amount of natural noise.

Our own previous work [8] aimed at mapping common recommender system methods - such as rating an item - to reasonable gesture and motion interaction patterns. We provided a minimum of two different input methods for each application function (e.g. rating an item). Thus, we were able to compare user interface options. We conducted a user study to find out which interaction patterns are preferred by users when given the choice. Our study showed that users preferred less complicated, easier to handle gestures over more complex ones.

Most of the existing studies do not take the mobile scenario into account, i.e. were not focussed on the interaction on mobile devices. When interacting with mobile devices, users may not be concentrated while being on the move or being distracted by the environment. Negulescu et al. [4] examined motion gestures in two specific distracted scenarios: in a walking scenario and in an eyes-free seated scenario. They showed that, despite somewhat lower throughput, it is beneficial to make use of motion gestures as a modality for distracted input on smartphones. Saffer [7] called these motion gestures free-form gestural interfaces which do not require the user to touch or handle them directly. Using these techniques the user input can be driven by the interaction with the space and can overcome some of the limitations of more classical interactions (e.g. via keyboards) on mobile devices [6].

In constrast to the existing work, we investigate the effect of user interaction methods on rating behavior on mobile devices (smartphones). We apply different input methods and interaction gestures in our interface to explore which ones decrease noise in the rating process. In the corresponding user study, we investigate the possible source of noise in rating results provoked by different input methods in the rating process. This study provides and analyzes the impacts of different interaction modalities on smartphones in the user giving feedback proceeding in details with the aim of overcoming rating result noise and enhancing recommender system quality.

## 3. INPUT METHODS IN THE TEST APPLICATION

To address this research question, we extend our previous work of a mobile recommendation application [8] . The scenario is a movie search and recommendation application that is similar the Internet Movie Database (IMDb) mobile application[1].

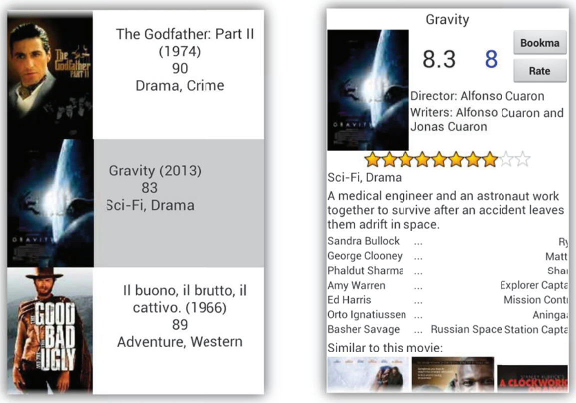
On the main screen, users can browse through the items to select a movie from the list (see Figure 1 (a)). Once they find a movie they are interested in, a single tap on that entry opens a new screen containing a more detailed description of the movie (Figure 1 (b)). Users can rate movies on a score from 1 (worst) to 10 (best) stars. To perfom the rating, they can choose one of the following three input methods:

1. On-screen button: users can rate a movie by selecting the "rate" on-screen button. The actual rating is performed by a simple tap on the 1 to 10 scale of stars (Figure 1 (b)).

2. Touch-screen gesture (*One-Finger_Hold_Pinch*) [8]: This rating uses a two-finger gesture. One finger is kept on the screen, while the second finger moves on the screen to increase or decrease or the rating stars respectively.

3. Free-form gesture (*Tilt*): Tilting means shifting the smartphone horizontally which is determined using it's gyroscope sensor. Shifting to the right increases the rating and shifting to the left decreases it. This rating is performed and saved without a single touch.

[1]see http://www.imdb.com/apps/?ref_=nb_app




(a) (b)

**Figure 1: (a) List of movies. (b) Details of movie screen.**

# 4. USER STUDY

## 4.1 Gesture Investigation

We conducted a user study to examine how a user's rating is influenced by the chosen input method. Another objective of this study was to evaluate the intuitiveness and efficiency of mapping input methods to some common recommender systems' functions in particular in a mobile scenario with a low attention span.

## 4.2 Procedure

At the beginning of each session, the task was explained to the users and the participants were asked to choose and rate 16 movies. The movies and corresponding ratings were recorded manually, not in the mobile application. Then, we handed the smartphones to the subjects and the users were asked to re-evaluate their intended rating for the same movies using the explained three input methods: on-screen button, touch-screen gesture (*One-Finger_Hold_Pinch*) and free-form gesture (*Tilt*) in two different scenarios. Participants had to rate four movies using each of the three different input methods, and then could freely choose a preferred method to rate another four items. Afterwards, the errors of users' in applying ratings were calculated based on their initial ratings.

The study investigated two scenarios. The first scenario was conducted while the user is sitting and thus can concentrate on the task. In the second scenario, the user is walking and thus not fully concentrated. We name these two scenarios *concentrated case* and *non-concentrated case*. Thus, each scenario consists of 16 ratings the subjects have to perform. Each rating process only takes a few seconds.

After having finished the experiment, the respondents were asked to fill out an online questionnaire. The questionnaire contained three main categories: prior knowledge, concentrated case (sitting scenario), non-concentrated case (walking scenario). For each part, we inquired the intuitiveness and user preference and also asked for the users' opinion on how much they thought the different interaction methods would affect their rating result. At the end, the interviewer asked the participants for suggestions of other gestures suiting the rating function better. The results of the evaluation are presented in the following section.

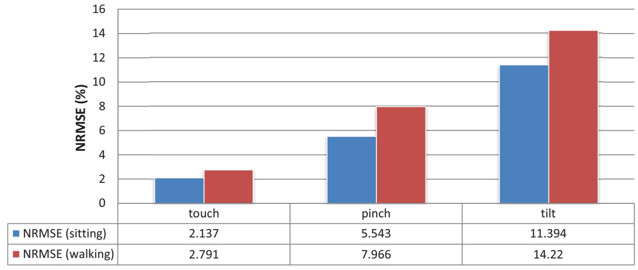



**Figure 2: NRMSE% for the three different interaction methods.**

## 4.3 Participants and Apparatus

20 persons participated in the study, mostly students and researchers of the Munich University of Technology. The experiment was performed using a Samsung Galaxy S III mini smartphone running Android 4.1.

# 5. RESULTS

## 5.1 Evaluation Methodology

We evaluate the error for every interaction method for rating by calculating the root mean squared error (RMSE) (formula 1). In formula (1), n equals to the number of rated movies, $\hat{y}_t$ denotes the user's intended rating, which was elicited before the beginning of the test application was started as mentioned in 4.2. $y_t$ is equivalent to the user's rating which was obtained from the test application log.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (\hat{y}_t - y_t)^2}{n}} \quad (1)$$

Figure 2 shows the normalized root mean squared error percentage (NRMSE%) which were derived from the following formula:

$$NRMSE\% = \frac{RMSE}{9} \times 100 \quad (2)$$

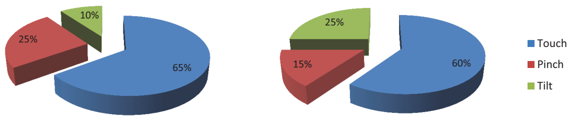
In this equation, 9 represents the maximum error since the values of ratings are in the range of [1;10].

## 5.2 Evaluation results

Figure 2 shows that the performance within the concentrated scenario is more precise than in the non-concentrated scenario, regardless which interaction method has been used. This was expected of course.

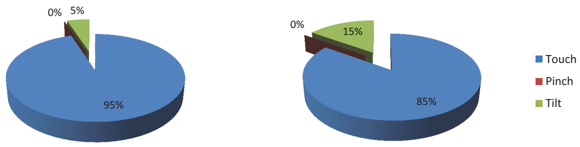
Among the different input methods, the on-screen button has the lowest error (with less than 3% very close to the intended rating), the touch-screen gesture has a medium accuracy, and the *Tilt* gesture has the highest NRMSE (more than 10%). Thus, the input method has a considerable effect on the resulting rating. In addition to that, the noise was lower towards the extreme ends of the rating scale.

The low score for *Tilt* might be caused by our implementation of the gesture and better calibration might change the result. However, less than ideal implementations of interaction methods may be present in many mobile applications.




(a) sitting scenario (b) walking scenario

**Figure 3: User preferred gesture.**




(a) sitting scenario (b) walking scenario

**Figure 4: Intuitivity of different gestures.**

At the end of each session, the participants were asked to rate four movies using the preferred interaction method which was logged afterwards. The goal of this part was to determine which input method is preferred depending on the specific scenario (sitting or walking). Figure 3 illustrates the results. Our subjects preferred the on-screen button as input method in both scenarios. However, *Tilt* and *One-Finger_Hold_Pinch* were assessed differently depending on the scenario. Participants preferred *Tilt* in the non-concentrated (walking) scenario over *One-Finger_Hold_Pinch*, but vice versa in the concentrated (sitting) case.

We also asked the participants how intuitive they found the three input methods for rating on a scale from 1 to 5 with 5 being "very intuitive". Figure 4 illustrates which methods were rated as more intuitive by the participants. The results show that the *on-screen button* was rated as most intuitive in both scenarios, while *Tilt* being the second highest but still with minor percantage in the walking scenario. This may be due to the fact that the on-screen buttons are commonly used in mobile applications and most people are used to it.

In our survey, we defined an intuitive gesture as "being easy to learn and a pleasure to use". There is a difference in what the users found intuitive and what they actually preferred. Our participants found the common and simple on-screen button as most intuitive but 35% preferred the other options in the sitting scenario and 40% in the walking scenario, respectively.

## 6. CONCLUSION

Customer trust is the critical success factor for recommender systems. Since recommender systems frequently depend on the users' ratings, there is a need to reduce the users' rating errors in order to improve the reliablility of recommendations. In this study, a new source of errors in the rating process on mobile phones was investigated. We showed that rating results differ depending on the interaction method. Thus they distort the actual rating of the user, which can be improved by using more intuitive and easy to perform gestures. In our study, the results of the on-screen button appear to be more precise and reliable being near to the user's stated actual opinion.

We also demonstrate that free-form gestures such as *Tilt* are somewhat more desired in non-concentrated scenarios. When the environment is distracting, free-form gestures are more embraced by users even though, as a nature of non-concentrated situation, the results contain some noise. Due to the mobile phone's character, users are willing to be able to exploit their smartphones in situations which need less attention to perform an action, such as rating. To satisfy this requirement, a free-form gesture is applied in order to facilitate actions on mobile phones.

Regarding future work, introducing and studying more free-form gestures is desirable for recommender systems especially in non-concentrated scenarios. Moreover, people may get more and more used to performing free-form gestures. Since the detailed implementation and calibration of free-form gestures may have effect, an optimized *Tilt* implementation may reduce the error for this input method, in comparision to the result in our study. Investigating voice input would also be an interesting research topic as they do not require much effort and attention.

## 7. REFERENCES


[1] Amatriain, X., Pujol, J.M., and Oliver, N. 2009. I like it... i like it not: Evaluating user ratings noise in recommender systems. In *Proc. of the 17th International Conference on User Modeling, Adaptation, and Personalization* (UMAP '09), 247–258. Springer.

[2] Cosley, D., Lam, S.K., Albert, I., Konstan, J.A., and Riedl, J. 2003. Is seeing believing?: how recommender system interfaces affect users' opinions. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '03). 585–592, ACM.

[3] Jawaheer, G., Weller, P., and Kostkova, P. 2014. Modeling user preferences in recommender systems: a classification framework for explicit and implicit user feedback In *Proc. of the 17th International Conference on User Modeling, Adaptation, and Personalization* (UMAP '09), 247–258. Springer.

[4] Negulescu, M., Ruiz, J., Li, Y., and Lank, E. 2012. Tap, swipe, or move: attentional demands for distracted smartphone input. In *Proc. of the Int. Working Conference on Advanced Visual Interfaces*, AVI '12, Capri Island, Italy, 173–180, ACM.

[5] Nguyen, T.T., Kluver, D., Wang, T.Y., Hui, P.M., Ekstrand, M.D., Willemsen, M.C., and Riedl, J. 2013. Rating support interfaces to improve user experience and recommender accuracy. In *Proc. of the 7th ACM Conference on Recommender Systems*, Hong Kong, China, ACM.

[6] Ricci, F. 2010. Mobile recommender systems. *J. of IT & Tourism*, 12, 3 (April 2010), 205–231.

[7] Saffer, D. 2008. *Designing Gestural Interfaces.* O'Reilly, Sebastopol.

[8] Woerndl, W., Weicker, J., Lamche, B. 2013. Selecting gestural user interaction patterns for recommender applications on smartphones. In *Proc. Decisions @ RecSys workshop*, 7th ACM Conference on Recommender Systems, Hong Kong, China, ACM.