Evaluating the Influence of Targets and Hand Postures on Touch-based Behavioural Biometrics

Daniel Buschek, Alexander De Luca, Florian Alt
University of Munich (LMU), Germany
{daniel.buschek, alexander.de.luca, florian.alt}@ifi.lmu.de

ABSTRACT
Users’ individual differences in their mobile touch behaviour can help to continuously verify identity and protect personal data. However, little is known about the influence of GUI elements and hand postures on such touch biometrics. Thus, we present a metric to measure the amount of user-revealing information that can be extracted from touch targeting interactions and apply it in eight targeting tasks with over 150,000 touches from 24 users in two sessions. We compare touch-to-target offset patterns for four target types and two hand postures. Our analyses reveal that small, compactly shaped targets near screen edges yield the most descriptive touch targeting patterns. Moreover, our results show that thumb touches are more individual than index finger ones. We conclude that touch-based user identification systems should analyse GUI layouts and infer hand postures. We also describe a framework to estimate the usefulness of GUIs for touch biometrics.

Author Keywords
Touch Targeting; Behavioural Biometrics; Mobile Device

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation (e.g. HCI): Input devices and strategies (e.g. mouse, touchscreen)

INTRODUCTION
Observing touch behaviour can yield important information about the user. This attractive data has been widely used in recent HCI research to personalise interfaces and to recognise individuals. Related work observed touches to tailor keyboards to the individual typist [3, 16, 20, 24, 45], to personalise fonts [14], to add an implicit layer to authentication (e.g. for unlock patterns [1, 18]), or to create new authentication methods based on touch biometrics [38]. Targeting personalisation, privacy and security, these applications demonstrate the usefulness and importance of extracting and recognising user-specific information in mobile touch behaviour.

Related work has mostly targeted specific applications of touch biometrics, for example pattern unlock screens [18]. Recently, features for touch biometrics have also been studied across a greater range of tasks [44]. Improving our knowledge of what to observe (i.e. touch features) is important. On the other hand, to develop robust touch-based behavioural biometrics, we also need to investigate what influences the degree to which users will exhibit individual touch behaviour. We also lack a formal metric for such individuality.

Hence, to facilitate research and applications that use touch biometrics, we contribute: 1) an approach for measuring user-revealing information in targeting behaviour; 2) insights into influences of interface targets and hand postures on this individuality; and 3) an evaluation framework to estimate the expected amount of user-revealing information in touch interactions with given interfaces. Our insights inform applications of touch biometrics – focus on characteristic interactions and ignore those revealed as less individual (Figure 1).

APPLICATION CONTEXT
To illustrate the kind of application we aim to facilitate with our analytical contributions in this paper, consider that a user accidentally leaves behind a phone in a bar after unlocking it. A stranger, or a nosey friend, picks it up to read emails, view pictures and so on, thus compromising privacy and security.

This threat can be addressed with continuous implicit authentication [44]. If the observed touch behaviour deviates from the legitimate user’s characteristic behaviour, the phone raises alarm, locks itself, or blocks access to sensitive data and apps.

For such a system, it is vital to know which touch interactions consistently yield characteristic information about the user – and which ones are generally not very individual. For example, we will see that large buttons lead to little individual information compared to smaller ones. Hence, the system described above could mitigate the risk of false user rejections and false attacker acceptances by taking into account that touches at large buttons should not be used to determine whether the legitimate user is interacting.
RELATED WORK
We relate our work to 1) research that shows why and how touch targeting behaviour varies between users; and to 2) systems which use such implicit individual touch information.

Variations in Touch Behaviour
Related research found sources for variations in touch behaviour and models to describe it, such as Fitts’ Law [21] and its refined version for touch input [9]. Furthermore, research showed that finger angles affect touch contact areas on tabletops [22, 40]; touch-to-target distances and directions depend on the finger’s pitch, roll and yaw, as well as head position [27]; fingers occlude the target [5], and users rely on different visual features of the finger tip to align it with the target [26].

Other related work examined how touch behaviour is influenced by target size [34], location [35], and hand postures [43], but without deriving individual user models. Targeting errors (i.e. offsets) were compensated with polynomial functions derived from touches of many users [25]. This was improved with flexible models trained on user-specific touch data for thumb input aiming at cross-hair targets [15, 41, 42]. In this paper, we also employ offset models, but for representing and recognising users touching at four different target shapes and with two hand postures.

In summary, we can expect individual touch behaviour due to variations in targeting, perception and anatomy. Without special hardware or expensive computations [37], features like finger angles or head locations remain hard to assess in practice. Hence, to develop an applicable measuring approach and evaluation framework, we consider such influences based on how they reflect in users’ targeting error patterns. These can be captured with offset models on any off-the-shelf mobile touchscreen device.

Utilising Individual Touch Behaviour
Many systems utilise individual touch behaviour: on-screen keyboards adapt to the typist’s touch behaviour to decrease error rates [2, 3, 45]; users can be identified by typing rhythm [32], and strokes on gesture keyboards [10]; pattern unlock can be enhanced with an implicit authentication layer using touch sequences [1, 7, 18]; other work proposed touch biometrics to replace passwords or patterns entirely [38].

Further work distinguished users on rear-projected tabletops based on typing touches [31] and fingerprints [28]. Identification with offset models has been tested for pairs of users with thumb input and cross-hairs [15]. In general, a wide range of touch features is used to infer or verify user identity: acceleration, pressure, size or timing [1, 7, 47], as well as gestures for zooming and scrolling [23, 46]. Xu et al. [44] have recently evaluated touch features for implicit authentication in four such tasks, including on-screen writing with the finger.

In conclusion, individual touch behaviour has been successfully utilised to improve usability and security of mobile interactions. However, behavioural observations were either interface-agnostic (e.g. scrolling [23, 46]) or tailored to specific interfaces (e.g. keyboard [2, 3, 45], unlock screens [1, 7, 18, 47]). These interfaces were mostly assumed to yield characteristic behaviour without further comparisons. We lack an understanding of the interface’s influence on user-specific touch information. To inform applications of touch biometrics, possibly across different apps and interfaces, we evaluate how interface targets and hand postures influence individuality of touch behaviour. Our insights allow applications of touch biometrics to optimise user observation schemes by focusing on the most individual interactions.

Recent work [8] has simulated touches to evaluate keyboards (e.g. error rates). We present a framework to simulate touch behaviour, not restricted to keyboards, to predict expected individuality of user behaviour with given interfaces. This helps applications of touch biometrics to gain useful expectations and to favour – during runtime or UI design – those interfaces that provoke highly individual touch behaviour.

MEASURING TOUCH INDIVIDUALITY
We present an information-theoretic measure for individuality of touch targeting behaviour, computing an “index of individuality” $I$. We define individuality as behavioural information (i.e. patterns), which is characteristic (varies between users) and consistent (stays similar for a user over time).

Rationale: Intuitively, our metric considers behaviour individual if it allows an observer to recognise individual users. Note that we do not aim to create a running identification application, but a concept to analyse collected touch targeting data. This only hypothetically includes recognising users to measure how characteristically and consistently they touch. More precisely, we compute an “index of individuality” $I$ (in bits) by measuring reduction in uncertainty about a user’s identity, achieved by analysing this user’s touch data.

Overview: The remainder of this section develops our measurement process. Figure 2 shows this, from left to right:

1. Observing touch targeting behaviour in a study with two sessions per user yields touch data for different tasks.
2. Gaussian Process regression applied to data from the first session captures users’ targeting error patterns. Figure 2 shows an example contour plot of a user’s horizontal errors.
3. Matching patterns with touch data from the second session results in probabilities indicating how likely each user is responsible for the analysed touches.
4. Individuality of user behaviour is quantified by the extent to which the uncertainty of user identity can be reduced with this extracted implicit user information.
Hence, the metric considers users’ individual characteristics (user-specific modelling in step two) and consistency (recognition of users’ patterns across two sessions in step three). We next describe modelling, matching, and measuring in detail.

**Representing Users**

To represent and recognise users’ individual targeting behaviour, we adopt touch-to-target offset models, originally proposed to correct sensed touches to improve accuracy [42].

**Model:** We use Gaussian Processes (GPs) [36] following Weir et al. [42], a method shown to capture user-specific touch behaviour. We refer to these sources for details on the model.

**Training:** We measure offset vectors between target centres and touch locations by asking users to touch targets on the screen. Assuming centres is no limitation: If users aim for a different point, offsets will simply be shifted with no change to the pattern. The model is trained on this data to predict offsets for touches (see [36, 42] for details on training GPs).

**Prediction:** Trained models predict a bivariate Gaussian of likely offsets, $N(\mu, \Sigma)$. To improve touch accuracy, we can add the mean prediction $\mu$ to the sensed location $t$ to correct it. The full predictive distribution of likely target locations is $N(t + \mu, \Sigma)$. We use these predictions to recognise users.

**Recognising Users**

To recognise users, we match observed offsets with offsets predicted by users’ models to derive the likelihood of each user being responsible for the analysed touches: We evaluate predicted by users’ models to derive the likelihood of each user being responsible for the analysed touches: We evaluate the observed behaviour. We thus interpret it as (Figure 3). This likelihood is high if the prediction matches the true target location under the GP’s predicted distribution. To recognise users, we match observed offsets with offsets likely (compare Eq. 2 and 3). Illustratively, if we “rule out” half of the users with the predictions, we save one bit to explicitly encode IDs of the remaining users ($\log_2 N - \log_2 \frac{N}{2} = \log_2 N$).

**Explicit Encoding User Identity**

In a theoretical view, explicitly encoding identity requires $\log_2 N$ bits per user for $N$ different users, since we need to represent numbers 0 to $N-1$ (hypothetical user IDs). This is the maximum entropy [30] of a uniform distribution over $N$ states (i.e. users):

$$H_{max} = -\sum_{i=1}^{N} \frac{1}{N} \log_2 \frac{1}{N} = \log_2 N \quad (2)$$

Entropy can be seen as a priori uncertainty [4]: Highest uncertainty ($H_{max}$) occurs when all users are equally likely (e.g. guessing). In other words, we need $H_{max}$ bits (of explicit information) to resolve this uncertainty about the user’s identity. We consider this the cost of explicitly encoding identity.

**Implicitly Inferring User Identity**

We consider a model (Eq. 1) that predicts the probability $p(u_i)$ that analysed touch data comes from user $u_i$. The entropy of the model’s predicted distribution over users is [30]:

$$H = -\sum_{i=1}^{N} p(u_i) \log_2 p(u_i) \quad (3)$$

Relying on such predictions, users are no longer equally likely (compare Eq. 2 and 3). Illustratively, if we “rule out” half of the users with the predictions, we save one bit to explicitly encode IDs of the remaining users ($\log_2 N - \log_2 \frac{N}{2} = 1$). We consider this the gains of implicit information.

**Individuality as Reduction in Uncertainty**

In summary, analysing behaviour reduces uncertainty about the user’s identity. It is this difference in uncertainty that we have gained as user-specific information. Hence, we measure the index of individuality $I$ via this reduction in uncertainty:

$$\text{Uncertainty Reduction: } I = H_{max} - H \quad (4)$$

If we interpret this equation as information gain (e.g. as in decision trees), it compares uncertainty before and after analysing user behaviour with the model. In information theory, it is also an equation for redundancy (see [30]).

To illustrate this interpretation, consider the extremes: $I = H_{max}$ if analysing behaviour removed all uncertainty ($H = 0$). If we are already certain about the user’s identity based on behaviour, the explicit identity seems redundant. However, if analysing behaviour cannot reduce uncertainty ($H = H_{max}$), we gained no implicit information and $I = 0$. Then, explicit identity is not redundant. Between these extremes lies a continuum where uncertainty is only partly reduced by analysing the user’s behaviour with the model.
Considering Correctness of Inferred Identity

While low entropy \( H \) indicates certainty, it does not imply correctness. For example, a model might predict user \( A \) with high certainty, while ground truth is user \( B \). Even if the model is entirely certain, if it is wrong, we cannot say that the explicit (true) identity is redundant information. Hence, we cannot reduce the costs of explicit identification, when a decision based on the model’s predicted probabilities would be wrong. Thus, we consider \( I = 0 \) in these cases:

\[
\text{Index of Individuality: } I = \begin{cases} 
H_{\text{max}} - H & \text{if } \max p(u_i) = u_t \\
0 & \text{otherwise}
\end{cases}
\]  

(5)

Here, \( u_t \) denotes the ground truth user. In this paper, we use this equation to quantify individual information in user behaviour. To estimate user probabilities \( p(u_i) \) needed for \( H \), we use our user recognition model defined in Equation 1.

Remarks on the Metric

In summary, we measure individuality as reduction of uncertainty about user identity, achieved by observing touch behaviour. Equation 5 can be related to common classification accuracy (ratio of correct user identifications). For discrete predictions (i.e. \( p(u_i) = 1 \) for exactly one user, and 0 for all others), accuracy is equivalent to the average value of \( I \), except for its scaling factor \( H_{\text{max}} \) and unit (bits).

In our probabilistic perspective, \( I \) is more informative than classification accuracy, since it measures reduction of uncertainty with correct predictions, instead of just counting them. Moreover, we consider bits more appropriate than accuracy for the amount of individual information in user behaviour. Our metric also respects the sample size of the analysed dataset (\( N \) in \( H_{\text{max}} \) term), whereas accuracy values always fall between 0 and 100\% regardless of \( N \).

We defined individuality as characteristic and consistent information in behaviour. We will see that our metric is sensitive to both: 1) It is sensitive to consistency, as shown by higher values when analysing behaviour within one session compared to analyses over a week. 2) It is sensitive to characteristic behaviour – obtained values reflect that behaviour characteristic among many users is to be considered more individual than if it was only characteristic among two users.

TARGETING STUDY

Touch targeting data was collected in a user study (see [11] for other analyses on this data, not related to individuality).

Study Design

The study followed a repeated measures design with independent variables hand posture (right thumb with device in right hand; right index finger with device in left hand), target location (400 locations), and target type: CROSS, KEY (4×7mm), APP (9×9mm), and FILL (height 9mm), shown in Figure 1a). The dependent variable was touch location. There were two sessions per participant with a gap of a week.

Hence, the dataset contains 24 subjects × 2 sessions × 2 hand postures × 4 target types × 400 target locations = 153,600 touches.

Apparatus

Data was collected with a custom Android app on a smartphone (Nexus 5). Per task, it displayed 400 targets, one at a time, with locations chosen from a grid in randomised order. Targets were always fully visible – their centres were placed at least half their width/height away from screen borders.

Participants

The study was completed by 24 participants (10 female, 3 left-handed, mean age 26), all students. Almost all participants stated in a questionnaire prior to the study that they used either hand to operate touchscreens, all including their right hand. Hence, using the right hand in the study was not an unusual task for the left-handed people. Hand sizes were measured from the index finger to the bottom of the palm (mean: 179mm; range: 159mm to 194mm) [11]. Participants were compensated with a €15 gift card for an online shop.

Procedure

Each participant was invited to two study sessions in the lab, with at least 6 days in between. A session consisted of a series of 8 targeting tasks (2 hand postures × 4 target types) and lasted for one hour, including a short questionnaire at the end.

Participants sat on a couch and held the phone in portrait orientation. They were asked to neither emphasise accuracy nor speed, but to tap “naturally”. A 8×8 latin-square was used for task order to minimise learning effects. Tasks alternated between thumb and index finger to reduce fatigue. Participants were encouraged to take breaks between tasks.

RESULTS

To evaluate individuality of touch targeting behaviour, we present four analyses on the described dataset: 1) we evaluate our method of representing user behaviour, 2) we measure individuality with the described approach, 3) we evaluate user differences in targeting features in more detail, and 4) we take a closer look at the role of the screen location of targets.

Evaluating the User Representation

This first analysis confirms that offset models trained per individual improve accuracy for all our tasks (related work studied different tasks/models [11, 42]). This shows that the models capture user-specific targeting behaviour and are thus suitable user representations for measuring individuality.

Touch Accuracy

Figure 4 compares touch-to-target-centre distances (i.e. offsets) per target type. GP models reduced offsets for all tasks, evaluated with ten-fold cross-validation per user and task. Hyperparameters (\( \gamma = 2, \alpha = 0.1, \sigma^2 = 0.001 \)) were chosen by cross-validation on one arbitrary session [42].

Greenhouse-Geisser corrected ANOVA found significant main effects of hand posture (\( F_{1.23} = 9.34, p < 0.01, \eta^2_G = 0.02 \)), target type (\( F_{1.10.25.21} = 104.38, p < 0.001, \eta^2_G = 0.44 \)) and model (\( F_{1.23} = 201.72, p < 0.001, \eta^2_G = 0.36 \)). The interaction effect of target type × model was significant (\( F_{1.01.23.22} = 163.15, p < 0.001, \eta^2_G = 0.55 \)). Bonferroni corrected paired t-tests showed significantly smaller RMSEs with the model than without (index/KEY, index/APP: \( p < 0.05 \); others: \( p < 0.01 \)). Models thus significantly improved touch accuracy.
Area Targets: Hit Ratios
To evaluate accuracy for area targets, we examined the ratio of hits (touch in target bounding box). Figure 5 shows that larger targets were hit more consistently, and that thumbs missed targets more often than index fingers.

Greenhouse-Geisser corrected ANOVA revealed significant main effects of posture ($F_{1,23} = 19.62, p < 0.001, \eta^2_p = 0.06$), target type ($F_{1,17.27.00} = 183.72, p < 0.001, \eta^2_p = 0.61$) and model ($F_{1,23} = 52.56, p < 0.001, \eta^2_p = 0.60$). All interaction effects were significant: posture \times target type ($F_{1,75.40.33} = 17.49, p < 0.001, \eta^2_p = 0.04$), posture \times model ($F_{1,23} = 21.36, p < 0.001, \eta^2_p = 0.01$), target type \times model ($F_{1,08.24.86} = 25.19, p < 0.001, \eta^2_p = 0.05$) and posture \times target type \times model ($F_{1,32.30.36} = 9.34, p < 0.01, \eta^2_p = 0.01$). Bonferroni corrected paired t-tests revealed significantly higher hit ratios with the model than without for index/KEY, thumb/KEY and thumb/APP (all $p < 0.05$). The model also improved hit ratios for the other tasks (not significant).

Applying Models Across Tasks
Improvements by models trained and tested on different tasks were not as large as with task-specific data. Models trained on FILL only reduced RMSEs for the other posture with FILL. Models trained on index finger tasks reduced RMSEs in all tasks, but thumb models only improved other thumb tasks (and FILL with index). We conclude that offset models are more robust across target types than postures.

Conclusions
The results show that GP offset models fit individual behaviour in all tasks, demonstrated by significant improvement in targeting accuracy and hit ratios. Hence, the models can represent individual users. Furthermore, our analyses revealed that thumb input was less accurate than index finger input, and that models trained with thumb data only reduced offsets for thumb tasks, but regardless of the target type. As an exception, models trained on FILL targets did not reduce offsets for other targets.

Individuality of Targeting Behaviour
We now apply the developed approach (Figure 2) to measure targeting individuality according to Equation 5.

This section reveals a ranking of targets and hand postures by individual information in resulting touch targeting behaviour. We also compare measuring individuality based on data of all users to the most basic case of comparing only two users at a time. Finally, we apply our user representation and recognition method (Equation 1) to user identification, reporting achieved accuracy and practical implications.

In general, we plot individuality over time: We update measured values after each touch, using all touches so far to infer user identity implicitly (see Equation 1). This matches practical applications of touch biometrics, which can also consider all available information to identify users (e.g. a continuous implicit authentication system).

Ranking of Targets and Hand Postures
Here, we measure the amount of individual information that can be extracted within sessions (how characteristic?) and across sessions (how characteristic and consistent?).

Within sessions: We trained offset models for all users on half of their data from the first/second session. With the resulting trained system (Equation 1), we measured the individual information in the other half of each user’s data from the same session, then averaged over all users per task (Figure 6a): FILL did not reveal much individual information (<0.5 bits), but other targets reached ≈ 3 to 4 bits after 100-200 touches. Thumb input with CROSS and KEY reached the highest amount of individual information (>4 bits).

Across sessions: We trained offset models on each user’s first session data, using the resulting system to measure individual information in each user’s second session data, again repeated for all combinations of users per task, then taking the average. Figure 6b) shows that about 1 to 1.5 bits less individual information was measured than in the within-session case, since behaviour varied over time. Thumb input resulted in 1.5 to 3 bits for all target types (but FILL). Interestingly, APP targets were more consistent than KEY targets here, as indicated by higher individuality across sessions. Table 1 summarises these results, also including numbers from the following comparison to measuring individuality on pairs of users.
### Table 1. Tasks ranked by individual information, based on the average value within and across session evaluations when measuring on the whole group of users. Smaller targets and thumb input led to more individual behaviour. **APP** was more consistent than **KEY** for thumbs (higher values across sessions in the group case).

<table>
<thead>
<tr>
<th>Rank</th>
<th>Target</th>
<th>Hand Posture</th>
<th>Individual Information (bits)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>group (24 users)</td>
</tr>
<tr>
<td>1</td>
<td>cross</td>
<td>thumb</td>
<td>3.45</td>
</tr>
<tr>
<td>2</td>
<td>key</td>
<td>thumb</td>
<td>3.45</td>
</tr>
<tr>
<td>3</td>
<td>app</td>
<td>thumb</td>
<td>2.70</td>
</tr>
<tr>
<td>4</td>
<td>cross</td>
<td>index</td>
<td>2.64</td>
</tr>
<tr>
<td>5</td>
<td>key</td>
<td>index</td>
<td>2.80</td>
</tr>
<tr>
<td>6</td>
<td>app</td>
<td>index</td>
<td>2.44</td>
</tr>
<tr>
<td>7</td>
<td>fill</td>
<td>index</td>
<td>0.41</td>
</tr>
<tr>
<td>8</td>
<td>fill</td>
<td>thumb</td>
<td>0.42</td>
</tr>
</tbody>
</table>

*Comparison to Pairs of Users*

If a user A tends to touch too far left, and B too far right, then these very simple patterns are individual. Such simple patterns will not be unique to a single user any more, if we observe many more users. Hence, behaviour is inherently more individual, if it is individual in a larger group of users. This analysis demonstrates that this is reflected by our metric.

We repeated the analysis both within and across sessions for all pairs of users to measure the most basic user recognition task (binary: “A vs B”). The ranking of targets and hand postures overall matched the previous results (see Table 1).

### Table 1. Tasks ranked by individual information, based on the average value within and across session evaluations when measuring on the whole group of users. Smaller targets and thumb input led to more individual behaviour. **APP** was more consistent than **KEY** for thumbs (higher values across sessions in the group case).

However, absolute values were different: Our metric showed the desired sensitivity to characteristic patterns – behaviour unique among many users was correctly considered more individual than behaviour unique among only two users.

### User Identification

Beyond their role as a measuring instrument, the methods for representing/recognising users with offset models (steps 2 and 3 in Figure 2) could also be applied to identify users as part of a continuous implicit authentication system. Hence, we evaluated user identification accuracy as well.

Accuracy is the number of correct user predictions, divided by the total number of predictions. For each group size $2 \leq N \leq 20$, we drew 100 random groups and computed average identification accuracy, splitting data for training and testing as in the previous analyses. Figure 7 shows the results. They match our expectations based on the preceding analyses and our metric: Users were more accurately identified based on behaviour observed with smaller targets and thumb input.

In general, accuracy was lower for larger groups. Consistent behaviour (i.e. within session) resulted in up to about 90% accuracy after 150 touches. Varying behaviour across sessions resulted in up to about 70% after 300 touches. We also computed F1-scores in the same way, which resulted in the same rankings of the tasks, with values for all but **FILL** tasks ranging from 0.70 to 0.89 within sessions for groups of size 20, and from 0.37 to 0.74 across sessions.

### Conclusions

Table 1 shows the ranking of targets and hand postures by individuality: thumb input is more individual than index finger, and **CROSS** and **KEY** overall yield more individual information than **APP**. Little to none user-revealing information can be extracted from touches aimed at **FILL** targets.
Figure 8. Pairwise user differences in offset vector lengths. Thumbs led to larger differences between users than index fingers. The biggest targets (FILL) showed the largest differences.

Figure 9. Pairwise user differences in offset vector angles. For all but FILL targets, the thumb led to smaller offset angle differences between users than the index finger.

**Behavioural Differences in Offset Angles**
We examined angles between offset vectors $\mathbf{o}_1, \mathbf{o}_2$ from two users at the same target location, using a shifted and inverted cosine similarity (to measure distance, not similarity):

$$d(\mathbf{o}_1, \mathbf{o}_2) = \frac{1}{2} - \frac{\mathbf{o}_1 \cdot \mathbf{o}_2}{\|\mathbf{o}_1\| \|\mathbf{o}_2\|}$$

For each target type and pair of users we computed this distance at each target location. Averaging over all locations yielded the user difference scores.

Figure 9 shows that angles varied less between users for thumbs than index fingers, likely due to anatomical constraints: a thumb’s angle relative to the device is more restricted than a free hand’s index finger.

In contrast to offset lengths, FILL resulted in the smallest angular user differences. The shape of these targets offered little space to vary the angle relative to the centre. In addition, users mostly touched on the right half of these targets due to the use of the right hand in the study. For the other target types, differences grew with increasing area ($\text{CROSS} < \text{KEY} < \text{APP}$). Hence, we conclude that users individually made different use of touch areas with respect to targeting angles.

**Conclusions**
These results show that thumb input leads to similar offset angles, but individual offset lengths. In contrast, index finger input leads to similar lengths, but individual angles. Target size and shape, as well as hand postures, influence the main features of individuality in touch targeting errors.

**Relevant Screen Locations**
We also analysed which parts of the screen are most important to describe users’ individual touch targeting error patterns, using the Relevance Vector Machine (RVM) algorithm [39]. Related work used this method for touch offset prediction and for training set analysis [41]. RVMs solve regression problems like GPs, but their predictions are based on a small subset of all training examples, the relevance vectors (RVs). In contrast, GPs use the full training set for predictions.

The RVs found by the algorithm can be interpreted as the most important touches for modelling user behaviour. Thus, screen regions with many relevant touches across users are more descriptive. Figure 10 shows the distribution of RVs per task for both screen dimensions ($x$, $y$).

Overall, most of the touches that the RVM found to be relevant for describing users’ targeting patterns were located near screen borders and corners (“peaks” in Figure 10 mostly near edges and corners). Hence, interface elements (and thus touches) at these locations are expected to yield more characteristic information. In contrast, fewer RVs were located in the centre of the screen.

These results are in line with the findings in related work on cross-hair targets [41]. Moreover, we found that the RVM used less training touches to describe index finger patterns compared to thumbs. This suggests that index finger input leads to less complex offset patterns.
INSIGHTS INTO TOUCH TARGETING

We summarise our key insights to inform research on touch modelling and implicit identification/authentication:

- **Target size and shape influence the main features of individuality in touch targeting errors:** Offset lengths and angles varied more between users for larger targets, but broad target shapes resulted in more similar offset angles.

- **Hand posture influences the main features of individuality in touch targeting errors:** Offset lengths varied more between users for thumbs than index fingers, as thumb size restricts the reachable area. However, these restrictions also resulted in less offset angle variations for thumbs.

- **Thumb input is less accurate and more individual than index finger input:** Offset lengths were smaller and hit ratios were higher with index fingers, but thumb input resulted in more characteristic and consistent offset patterns.

- **Offset modelling for accuracy improvement is more robust across target types than across hand postures:** Models trained with thumb data only reduced offsets for thumb tasks, but regardless of the target type. Training on the large FILL targets did not reduce offsets for other targets.

- **Different screen locations best characterise behaviour for different targeting tasks:** The most informative screen areas for modelling individual touch offset patterns varied between target types and hand postures, but screen edges and corners were in general more relevant than the centre.

- **Offset patterns are user-specific, but not for all targets:** Offsets are known to be user-specific [42], but we revealed that this is not true for all targets. Patterns are less characteristic and consistent if only one dimension matters (FILL).

- **Individual touch characteristics stay more consistent for small targets:** User-specific behaviour can be recognised a week after training, but with lower accuracy, especially for targets which allow for multiple targeting strategies (e.g. aiming at the centre or right half of broad targets).

IMPLICATIONS FOR APPLICATIONS

In summary, our results show that GUIs could optimise their expected gain in individual information per touch as follows:

- **Use interface elements at the screen borders:** Touches at these locations are more descriptive of users’ patterns.

- **Avoid unnecessarily large targets:** Users touch less consistently if they can choose from a large area to hit the target.

- **Avoid very elongated target shapes:** Such targets lead to less diverse offset angles, reducing pattern individuality.

- **Use thumb-friendly layouts:** User characteristics are more evident with thumb input than with the index finger.

We could try to improve implicit identification by designing interfaces with the above statements in mind, but usability considerations should obviously come first. Hence, these implications are mainly meant to help choose between usable designs, with optimisation for biometrics as a secondary goal. Moreover, instead of changing GUIs, developers can use our implications to optimise which touches to “watch out for”.

Coming back to our initial example, an implicit continuous authentication system could mitigate the risk of false decisions by weighting touches higher, which are aimed at small compact targets near screen borders. With approaches to infer hand postures [11, 29, 45], systems can put more weight on thumb than index finger touches to infer user identity (e.g. via weighted samples in SVMs, or a prior in Bayesian methods).

INTERFACE EVALUATION

To facilitate application and verification of our implications, we now present an interface evaluation method. It takes a GUI layout as its input and outputs a (rough) estimate of the amount of individual information that could be derived from touch targeting interactions with that GUI. Developers and researchers can use this method to computationally compare different GUIs with respect to their expected usefulness for touch-based user identification and authentication.

Rationale: Intuitively, we use the collected data to train touch models for different interactions (target types, hand postures). We then simulate touches based on these models. Finally, we measure individuality of the simulated interactions. These values can then be compared between interfaces. In summary, our method comprises of the following three steps:

1. **Creating a model:** For a given interface, we train an interaction model on the study data from the second sessions.

2. **Simulating touches:** We next sample from this model’s distribution to simulate touch interactions for this interface.

3. **Interface evaluation:** Simulated touches are fed into our measurement system, trained on the first sessions’ data.

In summary, we use our study observations to enable evaluation of target combinations/layouts not directly observed before. The approach considers consistency, since touches are generated by a model trained on users’ second sessions, and fed to a system trained on their first sessions. It also considers user characteristics, since it evaluates interfaces with our measuring system based on individual targeting patterns.

Applications and limitations: Predicted individuality is not meant to be interpreted absolutely, since simulated behaviour is less complex than real interactions. Nonetheless, we can compare predictions (e.g. higher individuality predicted for interface A than B). With this knowledge, developers and researchers can 1) alter interfaces iteratively or 2) weight observations accordingly, both to yield more user-revealing touch information. Figure 11 summarises our evaluation framework. The corresponding detailed formal description follows.

Touch Interaction Model

We consider an interface $E$ with a set of GUI elements $e \in E$, and a set of users $u \in U$. Our touch interaction model for $E$ is then defined by the following factorisation of the joint distribution of users $u$, elements $e$ and touches $t$:

$$p(t, e, u) = p(t|e, u)p(e)p(u)$$

(7)

The individual parts are explained in detail below.

Expected touch locations per GUI element: $p(t|e, u)$ describes the distribution of touches for interface element $e$ targeted by user $u$ as a bivariate Gaussian:
Figure 11. Graphical user interface evaluation framework. An interface is given as a set of elements (blue). Each element e is described by its bounding box, type (e.g. APP), and importance p(e). We train inverse offset models from collected touches to predict a distribution of likely touch locations \( p(t|e, u) \) for each target e and user u. Over all users and targets, this results in a touch distribution for the whole interface (orange): \( p(t|e, u)p(e)p(u) \). We sample touches (red) from this distribution by drawing a user from \( p(u) \), an element from \( p(e) \), and then a touch location from \( p(t|e, u) \). Finally, we feed these simulated touches t into our measurement system to predict individuality (green). Values can be compared between different interfaces, for example to evaluate which one has higher potential for behavioural touch biometrics.

\[
p(t|e, u) \sim \mathcal{N}(\mu_{e,u}, \Sigma_{e,u} + \varepsilon \Sigma_{n}) \tag{8}
\]

Here, \( \mu_{e,u} \) denotes the mean prediction of an inverse offset model for user u given location and type of element e. \( \Sigma_{e,u} \) is the covariance matrix of this predictive distribution. Inverse means that the model (here we again use the GP model) is trained to predict touch locations given a target location [12].

We add noise with zero mean and covariance \( \varepsilon \Sigma_{n} \), set proportional to the element’s width/height with scaling \( \varepsilon \), a hyperparameter of the model. This reflects that users may touch different parts of (larger) elements, as revealed by varying behaviour in our study, and also allows us to estimate behaviour for elements with sizes not exactly observed during training.

**Relative importances of interface elements:** \( p(e) \) models the elements’ relative importances. For example, when evaluating a keyboard we can use the relative frequency of the characters in a given language. A uniform distribution means that each element is equally important to the user.

**Relative user frequencies:** \( p(u) \) models how often users interact. We assume uniformity – each user is equally likely.

**Possible extensions:** Our model assumes fixed hand postures, since participants were told which postures to use. However, it can easily be adapted to freely chosen postures h, for example as \( p(t, e, u, h) = p(t|e, u, h)p(e)p(u)p(h) \), where \( p(h) \) is the observed relative posture frequency. We can also include terms like \( p(h|u) \) (different users favour different postures), or \( p(h|e) \) (different elements provoke different postures).

**Simulating Touch Interaction**

To simulate interactions for the given interface, we sample touches from the joint distribution \( p(t, e, u) \) as defined by our model: 1) we draw a user u from \( p(u) \); 2) we draw an element e from \( p(e) \); and 3) we draw a touch location t from \( p(t|e, u) \).

**Example Application**

Three example applications show that the model’s predictions match the study insights. Figure 12 presents a list, a homescreen, and a keyboard. The figure further shows each GUI’s touch interaction model (Equation 7, density shown via contours for thumb input) and predicted (pairwise) individual information. The method predicts the highest individuality for the keyboard, followed by the homescreen and finally the list. We discuss these results in the light of our previous insights:

**Keyboard:** Keys are smaller than app-icons and list entries; thus, users touch more consistently. Predicted high touch density (Figure 12c) also lies in the most descriptive screen regions (Figure 10), yielding more informative touches.

**Homescreen:** In contrast, the homescreen uses larger targets, including the elongated searchbar at the top. Such targets lead to less consistent behaviour. Moreover, the homescreen features targets in the less descriptive centre region of the screen.

**List:** Finally, the list yielded little individual information. It only features FILL targets, which the study revealed to lead to very inconsistent behaviour.

**Hand postures:** Our method predicted higher accuracy for thumbs than index fingers, matching our study findings.

In summary, we demonstrated that our method can rank interface layouts according to expected individuality in targeting behaviour. Predictions match expectations from our previous analyses. The model can thus utilise given touch data to estimate targeting behaviour for interface layouts. This complements our insights with an applicable evaluation tool.

**SUMMARY AND DISCUSSION**

**Individuality of targeting behaviour:** We analysed and ranked eight targeting tasks by individuality (Table 1). Using data from two sessions, our measuring approach considered both user characteristics and consistency. Small targets result in more individual touch behaviour. Targets and hand postures influence whether targeting error length or angle is the main feature of individuality in targeting behaviour. Moreover, thumb input is less accurate but more individual than index finger input. The most descriptive touch locations are near the screen corners and edges.

**Measuring individuality:** We presented an approach to measure the amount of individual information in touch targeting behaviour. Our analyses showed that the metric has desirable properties: 1) Comparison of evaluations within and across sessions (Figure 6) showed that the metric is sensitive to consistency. 2) It is easier to extract characteristic behaviour for two users than for 24 – the metric also proved to be sensitive to this complexity of individuality. We thus conclude that our measuring approach renders the amount of individual information in mobile touch targeting behaviour assessable.

**Interface evaluation:** We applied our findings in a probabilistic touch interaction framework for interface evaluation and discussed three example cases. The predicted ranking matched our expectations based on the manual analyses of the collected data. The framework can thus help to comparatively assess which interface leads to more individual information.

**Implications for Touch Biometric Systems**

**Observe offsets:** Our results show that touch offsets are useful to distinguish users and could thus complement previous feature sets (e.g. [44]).
Consider context: Hand postures and targets render some touches more “telling” than others. In particular, to distinguish users, we recommend to rely on thumb touches targeting small, compactly-shaped targets near screen edges.

Integrate with other systems: Influences of targets and hand postures imply that applications relying on touch biometrics can benefit from close integration with other systems. GUI targets are known to the OS and (touch-based) recognition systems could provide the current hand posture [11, 29, 45].

Use multi-model methods: Our results show that touch-based biometric systems could potentially improve accuracy with target-specific and posture-specific models instead of one-fits-all models. A promising approach is a hierarchy of models that increase in specificity, as successfully used in keyboard individualisation [45].

Opportunities for Future Applications
Holistic biometrics across GUIs: While most current systems focus on one GUI or app (e.g. keyboard [13], Android pattern [18]), our results help biometric systems to utilise touch observations across multiple GUIs, by weighting relative importances (e.g. keyboard touches are more important than list touches to infer user identity).

Applications with prior knowledge: Our insights inform default parameters for biometric systems, for example to treat touches on small targets and in screen corners as more reliable. This could be realised, for instance, via weighted samples in an SVM or via priors in a Bayesian approach.

LIMITATIONS
Influence of user representation: Measuring individual information I requires a model to recognise characteristic behaviour. Absolute values of I thus depend on the chosen model, as in other information measures of behaviour [33]. To handle model-dependency, we chose the well-researched offset model and evaluated its suitability as a user representation in all our tasks. In general, our approach can also be used with other models for other biometrics (e.g. based on gait).

Analyses: Future work could study further element properties (e.g. their visuals), landscape device orientation, and other touch interactions (e.g. scrolling), possibly in every-day use. While a controlled targeting study like ours is more abstract than real GUIs, this control is needed to examine the factors’ influences. Nevertheless, our tasks used common GUI element shapes/sizes and natural hand postures, and single taps are the most fundamental and common touch interaction.

Interface evaluation: We do not claim to predict exact absolute individual information with our interface evaluation framework, which would require analyses on more interfaces and testing to fine-tune the parameters. This is beyond the scope of this paper. However, here we have shown that the ranking of three typical mobile interfaces, as predicted by our method, matches expectations based on manual data analyses. While limited by available data (e.g. with respect to target types), our concept is flexible to work with any dataset of touches and targets. It can thus easily be adapted, for example to respect hand postures chosen by users.

CONCLUSION AND FUTURE WORK
We have evaluated individuality of mobile touch targeting behaviour, measuring how characteristically and consistently users target GUI elements with over 150,000 touches in eight tasks. We presented a metric, data analyses, and an interface evaluation framework. As a key insight, applications of touch biometrics (e.g. implicit continuous authentication) should consider hand postures and properties of GUI elements.

With this knowledge, biometric systems can 1) favour interfaces which yield characteristic and consistent touch information, and 2) optimise user observation schemes to appropriately focus on the most individual and thus prolific interactions. We regard these contributions as a fundamental step towards robust, holistic applications of mobile touch biometrics for interface personalisation and usable privacy and security.

In future work, we plan to build a continuous posture recogniser and user authentication system, with this paper’s results as prior knowledge of individuality of touch interactions.
REFERENCES


