# Usability of Biometric Iris-Capture Methods in Self-Service Applications

Yevgeniy B. Sirotin, Ph.D., Scitor an SAIC Company,
Jacob A. Hasselgren, Scitor an SAIC Company,
Arun Vemury, Department of Homeland Security Science and Technology Directorate (DHS S&T)

Biometric technologies are gaining wide acceptance as a means of securely establishing individual identity. Common biometric modalities include recognition of fingerprints, faces, and irises. However, increased adoption of biometric technologies poses a significant challenge to users within a public setting. Users generally have little prior experience using biometric devices and prior evaluations have not systematically evaluated device usability with this naïve population. We tested an array of biometric modalities and methods both staffed and unstaffed use-case scenarios with a representative naïve subject population. Here we report on the usability of three iris collection methods: user-controlled, operator-controlled, and mechanically-controlled. Biometric performance was strongly determined by subjects' ability to correctly position and gaze at the iris device. Performance was poorest for the user-controlled method. Usability issues are the main differentiators of iris collection methods for a naïve user population.

## INTRODUCTION

Interest in using biometric technologies to identify individuals at border crossings around the world is increasing (Muller, 2010). This increase stems in part from the fact that biometric technologies can significantly enhance security. Through successful automation and integration of these technologies into kiosks or gates, travelers can use biometrics to assert their identity without interacting with port staff.

Several countries have implemented biometric-dependent trusted traveler programs (New Zealand Immigration, 2015). In the United States, Global Entry and NEXUS programs are examples of this effort. Trusted-traveler programs reduce staffing demands by pre-screening travelers and then relying on self-service biometric technologies to confirm their identity at the port. Indeed, in the U.S., biometrics are rapidly becoming the go-to technology for more general identification of travelers at land and air ports through the use of self-service kiosks (CBP, 2016). The interest in biometrics is not limited to entry operations. Biometrics may facilitate identification of travelers leaving the port without requiring redeployment of significant staffing resources (DHS S&T, 2012).

More generally, biometrics is gaining momentum as a method of establishing identity without requiring the distribution of any identity tokens. There is significant interest in this regard within airports, where individuals must be identified repeatedly at different steps within the process (Ghee, 2015). Biometrics are also useful for access control and to prevent fraud in the financial industry. In sum, we may be close to a future where your face, iris, or fingerprint is your primary means of identification.

Yet, despite this possibility, biometric technologies at airports often find little interest from their target user populations (Oostveen, 2014). One explanation is that users are not aware of biometric identification systems. Indeed, until recently, few people have had extensive experience with biometrics (a trend that is changing since the inclusion of biometric authentication in personal electronics). However, an alternate explanation is that users are put off by the poor usability of some self-service biometric systems (Oostveen, et al., 2014).

Until recently, the use of self-service biometric systems at ports of entry has generally been voluntary, with self-service biometric systems available as an alternative to a traditional staffed process (DHS S&T, 2014). In the future, however, the self-service biometric process may become the standard, rather than an alternative. This makes it timely to understand the usability of different biometric modalities and methods to develop usability standards and performance benchmarks which may facilitate biometric system selection (NIST, 2008).

In particular, interest in iris biometrics has increased in both staffed and self-service applications due to high potential identification accuracy without required physical contact (NIST, 2013). Although iris biometrics can have high matching accuracy, they present challenges for naïve users (Sasse, 2007; NIST, 2008) stemming from two major sources.

First, the user must know how to position their head and eyes within the capture volume of the scanner using only the distal cues provided by the device. Executing such movement efficiently requires practice and learning (Wulf, 2007).

Second, most iris capture methods require the user to maintain lengthy eye fixation on the biometric camera. However, humans typically move their eyes three times per second (Martinez-Conde, et al., 2004), much faster than usually required for an acquisition.

Here we examine how iris acquisition performance is impacted by the usability of the iris collection methods tested in both staffed and unstaffed use-cases with naïve users.

*The research for this paper was fully funded by the Department of Homeland Security Science and Technology Directorate on contract number W911NF-13-D-0006-0003. The views presented here are those of the authors and do not represent those of the Department of Homeland Security or of the U.S. Government.*

### Study Objectives

We performed a large-scale study to evaluate the usability of different collection methods for multiple biometric modalities, including iris. Most prior work in the biometric field has been performed with a focus on access control applications with experienced users or applications with the availability of a device operator. Our work is different in that our test scenarios focus on naïve users within self-service environments (Sasse, 2007). Here we report on the usability of different methods of iris collection. Analyses of other biometric modalities may be published in subsequent reports. All devices included in the testing were technically comparable. However, some iris collection methods tested poorly. Specifically, the user-controlled iris collection methods, frequently used for access control, performed poorly relative to both operator-controlled and machine-controlled methods. We suggest that this low performance stems from fundamental difficulties in learning new movements for the user-controlled iris method.

## METHOD

### Overview

This study is part of a larger ongoing effort to evaluate biometric technologies for use in identifying international travelers within an airport environment. Both staffed air entry and self-service air exit scenarios were evaluated using a subject population representative of the international traveling public. The overarching study design combined multiple factors not presently examined. For this work, we focus on the different methods of iris capture in both air entry (staffed) and air exit (unstaffed) scenarios. We used a full-factorial design blocking on subject demographics: age, gender, and eyeshade. Each subject performed three runs of the full experiment, which included biometric capture in entry and exit scenarios, in that order. Experiments for entry and exit scenarios were balanced for first order carryover effects using a Latin square approach. The principal factor examined here is the method of iris biometric collection with output variables of collection method efficiency, effectiveness and user satisfaction (described below). Open coding of video collected during testing was used to characterize the usability issues associated with different collection methods.

### Subjects

All work with human subjects was reviewed and approved by an institutional review board. A total of 293 subjects were included in usability testing at a dedicated facility. Subject experimental groups were blocked based on age, gender, and eyeshade (Table 1).

Table 1. Demographic Distributions of Test Subjects

| Age Bin | *n* | Percent |
|---|---|---|
| 18-28 | 87 | 30% |
| 29-37 | 56 | 19% |
| 38-47 | 53 | 18% |
| 48-58 | 66 | 22.5% |
| 59+ | 31 | 10.5% |
| **Gender** | ***n*** | **Percent** |
| Male | 118 | 40% |
| Female | 175 | 60% |
| **Eyeshade*** | ***n*** | **Percent** |
| Light | 65 | 22% |
| Dark | 226 | 77% |
| Heterochromatic | 2 | 1% |

*Light: blue and green eye shades; Dark: brown and hazel; Heterochromatic: multiple shades within or between the two eyes

Different experimental groups tested different combinations of entry and exit iris methods. All subjects were briefed regarding their role in the evaluation. The three entry iris methods were each tested with one third of the subjects. The two exit iris methods were each tested with half of the subjects. Roughly equal numbers of subjects used each of the two exit iris methods after each of the three entry iris methods. Due to technical issues, timing and/or biometric data from 56 subjects was not analyzed. Four subjects did not complete m-SUS questionnaires. This reduced the total number of subjects available for various analyses.

### Use cases

All work was carried out at a dedicated test facility configured to mimic typical biometric use cases for international air travelers: air entry and air exit.

In both scenarios, care was taken to simulate the intended context of use. Subjects proceeded through the facility with carry-on baggage, recapitulating appropriate physical constraints. Subjects waited in queues, with opportunity to observe those ahead of them prior to initiating procedures. The initial transaction in each queue was performed by a trained test staff member so that proper use could be observed at least once by all subjects.

*Staffed Use Case (Entry Booth):*

This use case was designed to simulate existing inspection operations carried out for qualifying arriving travelers at U.S. airports (US-VISIT) (DHS S&T, 2012). Staffed booths simulated the physical interaction between travelers and biometric technologies. Subjects approached a staffed booth equipped with an iris biometric device. Subjects first presented an identifying token to the booth operator. The booth operator scanned the token and directed subjects to present their biometrics using the iris capture method designated for the experimental group (user-controlled, machine-controlled, or operator-controlled). The operator made up to three attempts to successfully capture iris images. Following a successful iris capture or three failed capture attempts, the subject was instructed to proceed.

*Unstaffed Use Case (Exit Gate):*

This use case was designed to simulate notional self-boarding air exit operations. Iris biometric devices and travel document scanners were integrated into realistic mechanically-operated gates. To proceed through the gate, subjects had to scan the travel document, which initiated iris capture for a 1:1 verification of identity. Subjects were given up to three attempts to submit biometric images for verification. To succeed, iris images needed to be both collected and matched to the travel document. Collection continued for up to three attempts. Subjects were given alternate instructions based on biometric match results. Subjects were not explicitly aware of final match performance. Failed collection or match attempts resulted in instructions to "try again". Successful match attempts resulted in instructions to "proceed" through the gate. Three failed collection or match attempts also resulted in the "proceed" instruction.

**Biometric Capture and Processing**

*Iris Capture Methods*

All biometric devices selected for usability testing were first evaluated for biometric image quality. All included devices met the specific technical criteria tested in a controlled environment with expert users. Included devices produced ISO/IEC 19794-6 standard conformant images, a true acceptance rate (TAR) threshold near 97%; and biometric capture times near or below 10 seconds. Included devices captured images under infra-red (IR) illumination within a capture volume (a region in 3D space). Capture volume size was a major differentiating factor between devices.

For the staffed entry use case, three methods of iris collection were tested: operator-controlled, user-controlled, and machine-controlled. For operator-controlled, a trained operator moved the iris capture device, like a photo camera, such that the subject's iris was within the capture volume, and attempted iris capture. For user-controlled, a trained operator delivered verbal instructions provided by the device manufacturer to subjects, explaining how they should position their own irises for capture within the fixed device's capture volume. For machine-controlled, a trained operator instructed participants to stand at a specified location and initiated iris capture on a device capable of mechanically adjusting its optics to bring the subjects' iris within the capture volume.

For the unstaffed exit gate, two methods of iris collection were tested: user-controlled and machine-controlled. For both methods, iris capture was initiated by subjects' actions at the gate. For user-controlled, subjects positioned in front of the iris device using manufacturer's written instructions. For machine-controlled, the device adjusted its optics to bring the subjects' iris within the capture volume.

Importantly, user-controlled iris devices at entry and exit had different manufacturers and distinct instructions for use: verbal and written, respectively. The machine-controlled iris device was the same at both entry and exit.

*Baseline Images*

High-quality ground-truth iris images were acquired within a light controlled environment for all subjects during study enrollment. To avoid biasing the subjects, ground-truth images were acquired using a contact-controlled iris method which differed significantly from the evaluated methods. Users placed a binocular-like device on their nasal bridge, which brought the irises within the capture volume. Biometric matching was performed against these images.

*Match Algorithms*

The same industry standard biometric match algorithm was used for all tested devices. The algorithm produced high (above 97%) iris match performance in pre-test device quality evaluations. Effects of match algorithms and biometric image processing are outside the scope of this report.

**Usability Metrics**

To measure baseline naïve performance, the usability metrics efficiency, effectiveness, and satisfaction were evaluated on the first run of the experiment and are defined as follows.

*Efficiency*

Efficiency was measured as the time taken to complete a biometric transaction, referred to as transaction time. We defined transaction time as the time taken by subjects to interact with a device, starting from the initial prompt to begin the interaction until final notification to proceed.

For entry, transactions started when the operator prompted the subject to provide biometrics. Transactions ended upon successful capture-match or after of three failed capture attempts. For exit, transactions started following the travel token scan. Transactions ended upon successful capture-match or after three failed capture-match attempts.

*Effectiveness*

Effectiveness was measured as a function of biometric accuracy (TAR). For this evaluation we defined a transactional TAR as the fraction of successful subject transactions. A successful transaction was one which resulted in successful biometric capture and verification of the subjects' identity by matching to available baseline images.

*Satisfaction*

We measured satisfaction with each method of iris collection using a system usability questionnaire modified to fit our use cases (m-SUS). The 10 questions making up the m-SUS were converted into a single satisfaction score computed for each subject and each encountered device.

*Statistical Analysis*

Statistical quantities and comparisons were calculated using custom scripts written in R (https://www.r-project.org/).

**Video Analysis**

During testing, video cameras monitored each entry and exit transaction from two complementary angles allowing detailed observation of subject behavior. Video segments associated with each transaction were excised for analysis.

Video clips were analyzed using an open coding approach focused on identifying salient differences in subject behavior between short successful transactions and long and unsuccessful transactions. Several rounds of review by two independent analysts achieved final consensus for the final coding schema. Schemas were refined to minimize the number of distinct categories for the observed usability issues with each collection method. Due to the large volume of video data, analysis was restricted to the slowest 35 transactions (across all runs) using the fastest 35 transactions for reference.

## RESULTS

Subjects completed the entry followed by the exit use case transactions, yielding efficiency and effectiveness data. Satisfaction was measured after each transaction using the m-SUS survey. The sections below describe the usability metrics and video analysis results for the tested use cases.

## Staffed Use Case

The staffed use case was designed to simulate processing at staffed booths similar to those used by Customs and Border Protection at U.S. airports. At the booth, an operator guided subjects through the process of iris capture. We tested three capture methods at entry: operator-controlled, mechanically-controlled and user-controlled. Testing showed that operator-controlled and mechanically-controlled capture methods performed comparably well. The user-controlled method was significantly worse by all metrics: efficiency (Table 2) effectiveness (Table 3), and satisfaction (Table 4).

Table 2. Efficiency of staffed iris capture methods

| Biometric Method | *n* | Transaction Time (s) mean | s.d. |
|---|---|---|---|
| Operator-controlled | 81 | 40.7 | 22.6 |
| Mechanically-controlled | 73 | 40.2 | 18.1 |
| User-controlled | 83 | 88.6* | 43.6 |

**Significantly different than mechanically-controlled (Wilcoxon rank sum test, p=2.2e-16) and Operator controlled (p=2.2e-16).*

Table 3. Effectiveness of staffed iris capture methods

| Biometric Method | *n* | TAR | 95% CI* |
|---|---|---|---|
| Operator-controlled | 81 | 100% | 96%-100% |
| Mechanically-controlled | 73 | 97% | 91%-99% |
| User-controlled | 83 | 72%+ | 62%-81% |

**Wilson score confidence interval.*
*+Significantly lower than operator-controlled (2-sample test for equality of proportions, p=1.0e-6) or mechanically-controlled (p=5.7e-5).*

Table 4. Satisfaction with staffed iris capture methods

| Biometric Method | *n* | m-SUS Score (%) mean | s.d. |
|---|---|---|---|
| Operator-controlled | 100 | 86.3 | 15.3 |
| Mechanically-controlled | 94 | 87.6 | 14.3 |
| User-controlled | 95 | 68.6* | 26.8 |

**Significantly different than mechanically-controlled (Wilcoxon rank sum test, p=3.2e-13) and operator- controlled (p=7.1e-11).*

The poor performance of user-controlled iris relative to other methods could be explained by specific usability issues identified through open coding analysis of recorded video. To identify the most significant issues, we focused our analysis on the slowest transactions observed during testing.

The major classes of usability issues identified in the staffed use-case were: eye, interference, and position issues (Table 5). Gaze issues included subjects blinking, squinting, and moving eyes away from the biometric capture device. Interference issues included occlusions of the eyes by headdresses, hair, or excessive makeup. Finally, position issues comprised errors in subjects positioning themselves for biometric collection including correct distance and head angle.

Table 5. Usability issues associated with slow transactions of staffed iris capture methods

| Biometric Method | TAR | Issue Category, *n* = 35 Gaze | Interference | Position |
|---|---|---|---|---|
| User-controlled | 14% | 23% | 11% | 100% |
| Operator-controlled | 100% | na* | 20% | na+ |
| Mechanically-controlled | 91% | 49% | 20% | 0% |

**Eye issues were not identified as contributing to slowing Operator-controlled iris collection.*
*+Subject position issues were not identified for Operator-controlled collection.*

Interestingly, the longest transactions with staffed user-controlled iris collection had a TAR rate of only 14% and all were associated with position issues. This suggests that incorrect positioning by subjects was more detrimental to performance than other issue categories.

For the mechanically-controlled method, no position issues were identified, however, many subjects failed to fixate the device while keeping their eyes fully open long enough for an image to be acquired. For this method, gaze issues appeared most associated with failure to collect biometrics.

Finally, operators never failed to acquire subjects' iris images and neither gaze or position issues were observed. Subjects seemed to understand intuitively when their picture was being taken, knew where to look, and positioned themselves appropriately when prompted by the operator.

Overall, partial occlusions of the eyes or face or heavy makeup did not associate as strongly with poor biometric collection failures as other issues.

## Unstaffed Use Case

The unstaffed use case simulated an unstaffed biometric verification process suitable for boarding at an international exit gate. Currently travelers scan their boarding pass prior to entering the aircraft. In this use-case, subjects asserted their identity by scanning a boarding pass and then verified their identity using one of two iris biometric collection methods: mechanically-controlled or user-controlled.

The efficiency (Table 6) of the different iris methods was comparable. However, the user-controlled method was significantly worse than mechanically-controlled with respect to effectiveness (Table 7) and satisfaction (Table 8).

Table 6. Efficiency of unstaffed iris capture methods

| Biometric Method | *n* | Transaction Time (s) mean | s.d. |
|---|---|---|---|
| User-controlled | 118 | 24.2* | 6.1 |
| Mechanically-controlled | 120 | 23.0 | 14.8 |

**Significantly different than mechanically controlled (Wilcoxon rank sum test, p=5.1e-6). Likely due to differences in shape of distributions.*

For both methods, satisfaction was lower in the unstaffed as compared to the staffed condition ($p$<0.02, Wilcoxon rank-sum test). The accuracy of the mechanically-controlled method in staffed and unstaffed use-cases was comparable. The accuracy of the user-controlled method was conspicuously lower in the unstaffed configuration ($p$<1.1e-10, 2-sample test for equality of proportions) suggesting that verbal operator instructions may be better than written signage for user-controlled iris.

Table 7.Effectiveness of unstaffed iris capture methods

| Biometric Method | *n* | TAR | 95% CI* |
|---|---|---|---|
| User-controlled | 118 | 25%+ | 18%-33% |
| Mechanically-controlled | 120 | 95% | 90%-98% |

*Wilson score confidence interval.
+Significantly lower than operator-controlled (2-sample test for equality of proportions, p=2.2e-16).

Table 8. Satisfaction of unstaffed iris capture methods

| Biometric Method | n | m-SUS Score mean | s.d. |
|---|---|---|---|
| User-controlled | 138 | 52.1* | 28.6 |
| Mechanically-controlled | 150 | 76.9 | 24.1 |

*Significantly different than mechanically-controlled (Wilcoxon rank sum test, p=1.8e-13).

Table 9. Usability issues associated with slow transactions of unstaffed iris capture methods

| Biometric Method | TAR | Issue Category, $n$ = 35 Gaze | Interference | Position |
|---|---|---|---|---|
| User-controlled | 20% | 6% | 26% | 100% |
| Mechanically-controlled | 66% | 40% | 34% | 14% |

Usability issues associated with the unstaffed use-case were similar to the staffed use-case (Table 9). For the unstaffed use-case, the primary gaze issue was uncertainty about where, and for how long, to fixate the iris device, likely because no operator was available to direct subjects' gaze. Interference and position issues were as previously described for the staffed process. The longest transactions with unstaffed user-controlled iris collection had a TAR rate of only 20% and, as for the staffed use-case, all were associated with position issues. No other class of issues was as clearly associated with failed transactions.

Gaze issues were observed for many of the slow mechanically-controlled iris collections. Without operator instructions, many subjects were confused regarding when, where and for how long to look.

Position issues did not contribute strongly to unstaffed mechanically-controlled iris collection failures. Unlike in the staffed configuration, operators were not available to remind subjects to remove their glasses prior to collection. Consequently, eye occlusions (especially glasses) were more frequent and served to prolong transactions or lead to failure.

## DISCUSSION

The main behavioral issues associated with slow iris collection, or failure of iris collection, were difficulties with positioning within the capture volume and failures to look properly at the iris device.

For both user-controlled and mechanically-controlled capture methods, effectiveness and satisfaction were highest in the staffed use-case. However, staffed transactions required more time to complete. The user-controlled capture method tested worse than all other methods due to its heavy reliance on correct positioning and appropriate gaze.

User-controlled iris devices require subjects to execute a series of carefully controlled movements. However, efficient execution of complex motor patterns requires learning (Wulf, 2007) and naïve users were unable reliably follow movement instructions delivered verbally or by written signage. The user-controlled method tested better in the staffed use-case, suggesting that verbal instructions were more effective. Still, performance was worse than other capture options.

All iris methods require subjects to look at the iris camera for an amount of time in excess of normal fixation durations. Slow and failed transactions were associated with subjects understanding of where, when, and for how long to look during collection. As shown here, these issues can be mitigated by appropriate instructions delivered by staff. However, for unstaffed use-cases, shorter capture times and appropriate signage may be effective.

Presently, most international travelers are only infrequent users of iris biometrics and many would encounter the technology for the first time. Thus, in the short term, any iris device implemented at the border would have to overcome the challenges associated with appropriate positioning and gaze direction. Of the tested methods, mechanically-controlled iris collection tested well in this context of use even without staff. This is consistent with this method's relaxed positioning requirements and relatively clear gaze instructions. Iris collection method performance may be improved by mitigating identified usability issues; however, selecting the best method for the intended use case may be more expedient.

## REFERENCES


CBP, 2016. *Automated Passport Control (APC).* [Online] Available at: http://www.cbp.gov/travel/us-citizens/automated-passport-control-apc

DHS S&T, 2012. *Analysis of Operational Processes and Requirements for Biometric Exit Pilots,* Washington, DC: DHS S&T.

DHS S&T, 2014. *Apex Air Entry and Exit Re-Engineering.* [Online] Available at: http://www.dhs.gov/sites/default/files/publications/Apex%20Air%20Entry%20and%20Exit%20Re-Engineering-AEER-508_0.pdf [Accessed September 2014].

Ghee, R., 2015. *Aruba Happy Flow lays foundations for european preclearance..* [Online] Available at: http://www.futuretravelexperience.com/2015/06/aruba-happy-flow-lays-foundations-for-european-preclearance/

Martinez-Conde, S., Macknik, S. L. & Hubel, D. H., 2004. The role of fixational eye movements in visual perception. *Nature Reviews Neuroscience,* pp. 229-240.

Muller, B. J., 2010. *Security, Risk and the Biometric State: Governing Borders and Bodies.* New York: Routledge.

New Zealand Immigration, 2015. *Biometric Data for UK Visas.* [Online] Available at: http://dol.govt.nz/immigration/knowledgebase/item/1254

NIST, 2008. *Usability & Biometrics: Ensuring Successful Biometric Systems,* Gaithersburg, MD: NIST.

NIST, 2013. *IREX IV: Part 1, Evaluation of Iris Identification Algorithms,* Gaithersburg, MD: NIST.

Oostveen, A.-M., 2014. *Non-use of Automated Border Control Systems: Identifying Reasons and Solutions.* Southport, UK, s.n.

Oostveen, A.-M., Kaufmann, M., Krempel, E. & Grasemann, G., 2014. *Automated Border Control: A Comparative Usability Study at Two European Airports.* Lisbon, Portugal, s.n.

Sasse, M. A., 2007. Red-Eye Blink, Bendy Shuffle, A User Experience of Biometric Airport Systems. *IEEE SECURITY & PRIVACY,* pp. 78-81.

U.S. Department of State, n.d. *Safety & Security of U.S. Borders: Biometrics.* [Online] Available at: https://travel.state.gov/content/visas/en/general/border-biometrics.html [Accessed January 2016].

Wulf, G., 2007. *Attention and Motor Skill Learning.* Las Vegas, NV: Human Kinetics Publishers.