A Framework for Biometric and Interaction Performance Assessment of Automated Border Control Processes

Joshua J. Robertson[1], Richard M. Guest[1], Stephen J. Elliott[2] and Kevin O'Connor[2]

***Abstract*—Automated Border Control (ABC) in airports and land crossings utilise automated technology to verify passenger identity claims. Accuracy, interaction stability, user error and the need for a harmonised approach to implementation is required. Two models proposed in this paper establish a global path through ABC processes. The first, the *Generic Model,* maps separately the enrolment and verification phases of an ABC scenario. This allows a standardisation of the process and an exploration of variances and similarities between configurations across implementations. The second, the *Identity Claim Process,* decomposes the verification phase of the Generic Model to an enhanced resolution of ABC implementations. Harnessing a Human-Biometric Sensor Interaction framework allows the identification and quantification of errors within the system's use, attributing these errors to either system performance or human interaction. Data from a live operational scenario is used to analyse behaviours, which aid in establishing what effect these have on system performance. Utilising the proposed method will aid already established methods in improving the performance assessment of a system. Through analysing interactions and possible behavioural scenarios from the live trial, it was observed that 30.96% of interactions included some major user error. Future development using our proposed framework will see technological advances for biometric systems that are able to categorise interaction errors and feedback appropriately.**

***Index Terms*— Biometrics, Border Control, Human Error, Human-Biometric Sensor Interaction (HBSI), User, Performance Assessment**

I. INTRODUCTION

Automated Border Control (ABC) can be defined as the use of automated or semi-automated systems, which can verify a traveller is crossing the border at a control point without the need for significant (or any) human intervention [1]. The system aims to authenticate the traveller's claim of identity using a combination of biometric data, tokens or permits. The system will also attempt to establish whether the traveller is the rightful owner of a document/token, query border control records and watch lists, and then determine eligibility for border crossing permission. The implementation must also guarantee border security, preventing multiple subject entries from a single transaction (so-called tailgating). There should also be a manual inspection route for travellers who were refused entry (either correctly or erroneously) [2].

While variation in legal frameworks and entry requirements may prevent global harmonisation; all ABC systems should aim to provide a user-friendly experience. Travellers will judge how usable an implementation is based on prior knowledge of systems and the success of previous interactions. Their perception of convenience, confidence and their (subjective) satisfaction of the overall use of the system should be considered when evaluating the efficiency and effectiveness of ABC systems. All self-service systems (such as ATMs) require the user to draw upon the previous experience of similar automated interactions. Contrary to many other self-service systems, ABC systems typically are encountered less frequently. On average, an individual only travels abroad once or twice a year [3] which means experience and knowledge may not be adequate to ensure a smooth process through the system on future journeys. For the casual flyer, issues may arise due to unfamiliarity with the system which can be influenced by cultural, language and other ergonomic factors.

To evaluate the performance of ABC systems, we propose two models to identify process flow. A route-map of system components, modalities and requirements can be established for an implementation. We then utilise the Human-Biometric Sensor Interaction (HBSI) evaluation methodology to attribute system performance to user interaction and technologically based errors.

The contribution of these proposed methods will enable the design and research of ABC implementations to identify performance-related issues throughout the product life cycle. This research will establish the process flow of a border control system, identifying each step of the process involved and, therefore, enable the formal mapping of systems worldwide. The novelty of this research lies in the attribution of the HBSI framework to an identity claim scenario, which will allow the categorisation of user presentations made to a sensor. The limited research in biometric performance assessment has highlighted a need for a precise method of analysing these implementations beyond a system level. Applying HBSI will take the first steps to providing a full range of performance metrics which will ultimately improve the precision of biometric testing and reporting.

1.1 Automated Border Control

At a verification stage/arrival at the border, an ABC scenario will require a traveller to undergo identity verification through

[1] J.J.Robertson and R.M.Guest are with the School of Engineering & Digital Arts at the University of Kent, Canterbury, United Kingdom, CT2 7NT (e-mail: {j.j.robertson, r.m.guest}@kent.ac.uk)

[2] S. J. Elliott and K.O'Connor are with the International Center for Biometric Research at Purdue University, West Lafayette, IN 47907 USA (e-mail: {Elliott, koconnor @purdue.edu).

a series of interactions. Upon entering the system, the user will initiate an interaction through the use of an electronic travel document (such as passports or identity cards), commonly known as a 'token'. The token contains, or make reference to, an enrolled biometric sample against which a verification sample is compared. Typically the biometrics employ facial or fingerprint technologies [4]. Tokens are authenticated and checked for fraudulence [5]. Upon successful verification, the system will allow border crossing usually through the opening of a gate. Regarding topology, Frontex [2] classifies current ABC systems into three categories:

- One-Step Process: when the token verification, identity verification and the border crossing happens in one single process.
- Integrated Two-Step Process: when the token verification and eligibility to use the system is performed in advance and, if successful, the identity verification process is conducted at a different stage in the same physical location.
- Segregated Two-Step Process: when the process of traveller verification and the border crossing are completely separated. A further token is sometimes required to link both processes, sometimes in the form of a biometric sample or ticket.

Systems typically use physical barriers, full page token readers, visual displays for instructions, biometric capture devices and system management hardware and software. The systems may also include uniqueness and liveness detection technologies [6].

1.2 Interaction Aspects

Performance assessment concerning the interaction with devices (including biometric systems) is assessed from either a user perspective or by the effect on system performance through incorrect interaction. The usability community will assess the efficiency, effectiveness and user satisfaction [7] of systems from a user point of view. Other methods aim to provide valuable input into the design of systems through ergonomics, instructions and feedback. The work was undertaken in this current study purposefully takes a system perspective, in that we aim to establish the effect on the performance of a system when an erroneous interaction occurs.

The term usability is defined by ISO 9241-11 [7] by the extent to which a product, biometric or otherwise, can be used by subjects to achieve their goals. It can be assessed according to three criteria: efficiency, effectiveness and user satisfaction. Regarding an ABC system, it is possible to define task performance as effective when an interaction supports users who can achieve their goal of successfully crossing a border (including the sub-tasks of token reading and biometric verification). The interaction with the system is considered efficient if the traveller can pass through the process promptly, which is subjective to an individual user but averages at around 15-20 seconds for European ABC configurations [8] – [2]. A user's (subjective) satisfaction can depend on the level of the physical or mental workload that they may encounter throughout the process.

Research in the area of usability evaluation has been led by National Institute of Standards and Technology (NIST), who have contributed significantly to studying the assessment of usability in biometric systems [9] [10] [11] [12]. Other studies [13] [14] [15] [16] have investigated the influence of usability factors that affect biometric performance and user experience. This research demonstrates that while there are many parameters to consider throughout usability evaluations, these conditions may affect the user presentation at an interaction level.

Biometric components in border control solutions can cause problems. In some cases, the sample quality captured by the biometric component is insufficient, resulting in genuine token holders being denied access. In other instances, travellers found a particular modality awkward and time-consuming to use (as documented by a user experience study on the now retired UK IRIS programme [17] and the challenges of iris recognition in UAE [18].) Furthermore, other systems exhibited issues of inaccessibility resulting in a proportion of the population unable to use an implementation being unacceptably large [5] – [19]. A study on multiple verification systems conducted by the UK Passport Service also revealed some usability issues which affected system performance at an interaction level [20].

An important issue for travellers using ABC would be the system's ability to be able to communicate with people regardless of native language. Implementations that utilise a Segregated Two-Step Process with an interactive kiosk have an easier task of deploying (limited) language options [21], while one-step solutions offer little to no choice [2]. These configurations often rely on icons or simple pictorial instructions. If the user has previously experienced 'slow' system performance or has erroneously been denied access, these negative experiences may cause the traveller to avoid the process in the future [17]. How the system experience is conveyed through publicity documentation and to the public through the news media can also affect the user presentation [22]. A positive user experience is usually based on convenience, confidence that the system is functioning correctly, and its perceived utility [23].

There are also questions of user acceptance within biometric systems. The British Standards Institute (BSI) [19] found although most participants rated four systems they tested either satisfactory or positive; many raised several usability and acceptance issues. For example, within a fingerprint system, subjects commented on hygiene and the visible dirt which was highlighted due to illumination on the sensor. Current global consortiums such as the FastPass [24] and ABC4EU [25] projects have noted the need to find, standardise and counter non-technical factors. These often result in sporadic behaviour such as general confusion and unfamiliarity in different systems.

To enhance acceptance and to improve the user experience, an ABC implementation needs to accommodate: a population with different demographics, language barriers and travellers from a variety of cultural backgrounds through the standardisation of signage and instructions. Also, to travellers whose interactions may be affected by stress, fatigue and a reaction to unfamiliar surroundings. Furthermore, a system must exhibit an ability to convey errors and to offer solutions leading to a more efficient process for all travellers. Moreover, this must be able to accommodate user performance and acceptance concerns, accounting for confidence, and physical or mental workload.

1.3 Human-Biometric Sensor Interaction (HBSI)

In traditional testing, system metrics such as Failure to Acquire (FTA), a measurement of the percentage of invalid presentations that are incorrectly accepted as valid, can indicate issues of user performance [26]. Furthermore, Failure to Enrol (FTE), the rate at which attempts to create a template from an input are unsuccessful, will point to the success of individuals to interact with a system. Conventionally the overall performance of all biometric systems, including ABC implementations, monitor two key rates:

False Rejection Rate (FRR) - the percentage of incorrect rejections made by a system. False Acceptance Rate (FAR) - the measure of performance that a biometric system will incorrectly accept an access attempt by a non-authorised user.

These statistics, however, do not attribute the cause of the error, merely wrapping user and system performance within an individual metric. Performance results have been reported for several EU-based ABC implementations [27] which identified that FRRs were found to differ greatly between separate deployments.

Erroneous presentations which may be caused by unwanted user behaviour may result in rejection of the system. Correct presentations are rewarded with swifter processing times and a faster overall process. If incorrect user presentations could be categorised throughout the capture process, then appropriate algorithmic adjustments can be made. Development in this area could monitor and provide feedback to the user in an attempt to reduce overall error ratings. Eye tracking or image processing elements may be useful to identify when a user is making an incorrect behaviour, e.g. not looking at the camera. Travellers may often be tired or stressed when interacting with ABC systems, which may also have an effect on the process. Previous HBSI research has discussed the effect of human interaction on the biometric system [28], identifying that there is a relationship between user presentation and system performance.

The HBSI Evaluation Method presents an assessment method for determining systems' performance using multiple facets, technologies and interactions [29]. By doing so, this enables an investigation of the effect of user interaction when accounting for overall system performance. Ignoring these issues may cause the system not to achieve optimal performance, causing errors such as FTA and FTE and impacting the FRR.

Therefore, HBSI also aims to characterise the interaction between multiple components; the human and the sensor (ergonomics), the human and the biometric system (usability), and the sensor to the biometric system (sample quality). When reporting on usability, the interaction between the human and the biometric system should be described. Users will engage with a biometric system beyond interacting with multiple sensors. In ABC systems, for example, other components of the system such as the gate mechanisms and the feedback displayed on the monitor must also be considered. Regarding an individual interaction made to a biometric sensor, the HBSI uses the Presentation Framework (Figure 1) [30] – [31].

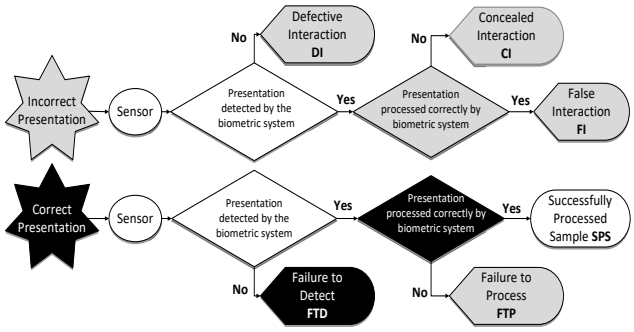

Fig. 1 The HBSI Evaluation Method

This process allows for an understanding of correct and incorrect behaviours typically occurring within a biometric system. Correct presentations for a particular interaction can be categorised as either Failure to Detect (FTD), Failure to Process (FTP) or a Successfully Processed Sample (SPS). SPS is the 'correct' transaction which is the ultimate goal of using a system. FTDs are correct presentations that are not detected by the system. FTPs within biometric systems can occur due to reasons such as problems in segmentation, feature extraction or quality control and is a system error generated by the biometric system.

There are three possible categorisations of incorrect presentations. Defective Interactions (DI) which occur when a biometric sample is incorrectly presented and is not detected by the system. Concealed Interactions (CI) occur when an incorrect presentation is detected by the system but is not handled correctly as an error. An example could be in fingerprint recognition where a user, for whatever reason, uses a different finger from that of the enrolled one but is still accepted by the system. False Interactions (FI) occur when a user presents their biometric in an incorrect way, and the system correctly identifies the error as an incorrect presentation. Recent work on the HBSI Model has investigated token presentations made to a sensor, creating a process chart that allows the categorisation of False Claims and Potential Attacks [32].

To enable an evaluation of ABC system performance, we propose the use of a generic method of the process flow. Onto this model, we can establish where HBSI errors occur at either a system or user interaction level and propose metrics for evaluating performance assessment.

The proposed model assess system performance at a task level. Our research is partially built upon the success of task analysis utilised in usability evaluations [33]. The tool is used to identify usability concerns for individual steps throughout a process. The success of our proposed model will take on a systems approach to understanding interactions. This will allow system administrators to identify bottlenecks in implementations that can be attributable to either ICT-based algorithmic performance or user interaction errors.

In addition to studying HBSI at a task level, usability metrics can also be established through performance assessment and post-usage questionnaires. The results can then be referenced back to the HSBI errors within a transaction. Deviation from a generic model may give indicators to user performance being affected by a new transaction sequence.

Section II describes a Generic Model (GM) for ABC enrolment and subsequent verification stages. Section III defines the outcomes within the verification process defined in

the GM. The Identity Claim Process (ICP) utilises the HBSI framework to identify possible system and interaction errors. The use of these two models is illustrated in Section IV using data from a live ABC implementation.

II. Generic Model

ABC systems across the globe use a broad range of biometric devices combining either single or multiple sensors and token readers. Requirements differ from country to country and have different usage implications for travellers depending on the configuration. To facilitate the application of the HBSI Presentation Framework to ABC, we first define a GM of existing systems, encapsulating key points across implementations. Performing cross-implementation comparisons are possible.

To encourage the development of our GM, we have analysed 21 global ABC implementations including the eGates from the UK as well as various EU, US, Australia, Singapore and Hong Kong systems. Out of the 21 systems, two configurations used a segregated two-step process while the remaining 19 were combined one-step solutions. 16 systems used a single modality for biometric verification (six using facial verification, five iris, four fingerprints, and one hand geometry) while five used multi-modal (all using fingerprints and face) technologies.

Three configurations were pre-registered systems and did not require a token. 11 used ePassport as the required token. The remaining seven used a combination of electronic IDs and electronically registered travellers' programme cards. All configurations involve a level of enrolment. However, pre-registered programmes using iris or fingerprint modalities required travellers to provide data at enrolment centres.

Although deviations do exist across implementations, a general process flow can be seen in the enrolment and verification stages. Systems are comprised of both automated (using technologies that do not require intervention by human operation) and manual elements. A GM of these automated and manual sections are shown in Figures 2 and 3. In our GM, grey sections refer to areas where manual intervention is required, while processes notated by a white section are automated (e.g. a biometric capture is algorithmically assessed, or a component can automatically detect movements within a gate). The white node indicates the starting point for interaction. Exit points within the GM, where travellers may be rejected from the system, are shown in black while white nodes with an outline denote success or approval through a process. Grey nodes refer to processes where a border guard may need to assist if the traveller is having difficulty with a certain action. For our definition of GM, we have constrained our route map to the major biometric modalities found in ABC systems.

The first stage of the GM will require the enrolment of biometric data to generate a token. After being approved onto a border access programme (E1), the traveller may be asked to provide biometric data at an enrolment centre or via self-captured samples (E2). Enrolment differs from each configuration reflecting the specific requirements and modalities.

TABLE I

Evaluation Points, Outcomes and Actions for Verification Stage

Evaluation Point	Definition	Possible Outcome	HBSI
V1 Traveller Presence	Is the traveller's presence detected?	Yes (A2), No (Reject/Assist)	FTD/DI
V2 Token Presence	Is the token detected?	Yes (A3), No (Reject/Assist)	FTD/DI
V3 Token Read	Was the token successfully read?	Yes (A4), No (Reject/Assist)	SPS, FTP/CI
V4 Biometric Capture	What biometric data is required?	Identify Modality (Iris, Finger, Faces)	All
V5 Data Verification	At what point does identification take place?	Database or Local level	N/A

It also permits the identification of the automated steps within each system, and where possible, the ability to identify errors which may occur throughout the presentation stages. HBSI applies to all outcomes of point evaluation E3 ('Biometric Data Capture') but is not applicable to E1 ('Traveller Eligible?') or E2 ('Biometric Data Required?'). Our GM includes provision for the detection of a traveller's document (V2) and thus the Presentation Framework can also be used to identify successful or unsuccessful token presentations. If the user had successfully entered their token, reading (V3) is performed through a sub-system process, and the result is fed back to the user. Upon successful validation of the token (if appropriate), subsequent biometric capture (A4) and verification at a local (token) or non-local (database) level (V5) of the traveller will lead to either authorisation or rejection to cross the border.

The Verification Stage (Figure 3) begins as a passenger enters an ABC system. Configurations may contain liveness detection components which can identify a passenger's presence (V1). The detection of the traveller can be assessed through an adapted version of the HBSI Presentation Framework. Whereby identifying conditions such as; whenever a user has entered the system correctly, too quickly, if another passenger is detected, or if the traveller is using the system already. Detection of such conditions are vital in the first stages of the process as this may alter how the system proceeds.

Table 1 shows the evaluation points throughout the Verification Stage of the GM, highlighting possible outcomes and HBSI categorisations. Although there are obvious points at which HBSI can be applied to improve the categorisation of errors, there are further steps where we can examine the GM further to enable an understanding of behaviours at presentation.

The verification stage of the GM can be decomposed further into individual interaction outcomes, mapping a clear process of interaction behaviours and systems steps throughout the ABC verification process. Doing so then allows the application of the HBSI Presentation Framework to specific steps and therefore, the user and system performance can be assessed through understanding and categorising traveller behaviours into scenarios.

TABLE II
IDENTITY CLAIM PROCESS AND RELATING GENERIC MODEL EVALUATION POINTS

Step	Title	Definition	GM
1	System Requires a Claim of Identity	The system may or may not require the user to make an identity claim.	**V1**
2	User Correctly Makes Identity Claim	The user either presents their token or submit their travel documents to the reader. The user must submit their token in such a way that the system should be expected to accept it.	**V2**
3	Identity Claim Accepted by System	If the token can be read then, it should be accepted by the system. If this step fails, it is a failure of the token or the system, not the user.	**V3**
4	Identity Claim Corresponds to Valid Identity	The token exists in the database, or the token has a valid enrolment sample, digital signature, expiry date. The token has not been revoked	**V3**
5	Claimed Identity belongs to a different user	The user may be using a false identity; for example, the token may have been (accidentally) swapped with a friend or travel companion. If the intent was malicious, then this counts as an attack.	**V3**
6	User Correctly Presents Biometric to System	A correct presentation can be defined when the user presents their biometric corresponding to the requirements of the system. It also means that they should submit the correct biometric trait e.g. the correct finger, iris which the system expects.	**V4**
7	Biometric Subsystem Detects Presentation	The biometric system correctly detects the biometric data and can perform subsequent processing.	**V4**
8	Biometric Subsystem determines that presentation is suitable for biometric matching	Biometric subsystem determines that the quality of the biometric sample be sufficient and can extract features to enable biometric matching to take place.	**V4**
9	Biometric matching validates user against claimed identity	If the system is an identification system, then this means that the user is determined to be an enrolled user. If it is a verification system, then the identity claim of the user is verified.	**V5**

III. IDENTITY CLAIM PROCESS

There are nine proposed steps for an Identity Claim Process (ICP) which occur throughout the Verification Stage as suggested in the GM. In Table 2, a definition of each proposed step and the related evaluation points from the GM is detailed. For example, Step 1 (System Requires a Claim of Identity) occurs during evaluation point V1. When a traveller enters the ABC interaction area, we can categorise if the system detects the traveller and if they are required to make an identity claim. Another example is at Step 3 (Identity Claim Accepted by System) which can only occur after V3 Token Read has been successful.

To facilitate our work, we begin to categorise user interaction within the ICP. While many of these scenarios are directly attributed to the user's presentation, there can also be algorithmic faults within a system sub-process that can lead to a particular error. Table 3 illustrates the possible outcomes for a scenario where the user has already entered the ABC system. In this example, the user has successfully had his or her token read (Steps 3 and 4) but has failed the step where they are required to present correctly (Step 6) to the sensor. In this situation, various outcomes can be attributed through HBSI categorisations but may be perceived differently in the system response.

In the illustrative scenario, only the system categorisation would be in effect in conventional assessment metrics. The inclusion of HBSI can help to establish cases where the system was correct in the identification of the scenario as erroneous (False Interaction). Therefore, we can indicate correct system performance or highlight potential security threats where the biometric is not detected, but the system grants access (Concealed Interaction).

Upon identifying these scenarios, the performance of recognition algorithms, human-computer interfaces and the ergonomics of the systems can be analysed in further detail.

IV. CASE STUDY OF ABC PERFORMANCE ASSESSMENT

We present a case study to illustrate the application of the HBSI method to the analysis of performance. Using the SmartGate/SmartGate+ configuration based in Australia and New Zealand, we apply both models to identify the requirements and differences between the original setup and the next generation of SmartGate (denoted as 'SmartGate+').

The travellers were observed entering the system from both sides of the gate through live video footage. A clear view of the entire transaction could be seen from both feeds. The footage was recorded over a period of two days.

4.1 SmartGate and SmartGate+

The original SmartGate system was an Australian and New Zealand airport ABC implementation employing a segregated two-step configuration. Upon arrival, the traveller approaches a standalone kiosk and enters his or her electronic passport to be read. An electronic ticket solely used for the gate interaction is then issued to the passenger. At the second stage, the passenger inserts his or her electronic ticket into a gate reader, after which a biometric facial verification subsystem matches the traveller's live photo with the reference image read from his or her passport. In 2013, SmartGate extended its services to UK and US citizens [34].

TABLE III
AN EXAMPLE OF AN IDENTITY CLAIM SCENARIO

ICP Step									System Categorisation	HBSI Categorisation
1	2	3	4	5	6	7	8	9		
Y	Y	Y	Y	N	N	N	Y	Y	True Match	Concealed Interaction
Y	Y	Y	Y	N	N	Y	Y	N	False Non-Match (User Fault)	Concealed Interaction
Y	Y	Y	Y	N	N	Y	N	N/A	Failure To Process	False Interaction
Y	Y	Y	Y	N	N	N	N/A	N/A	Biometric Not Presented	Defective Interaction

TABLE III
AN EXAMPLE OF AN IDENTITY CLAIM SCENARIO

GM	ICP STEP	SMARTGATE	SMARTGATE+
V1	1	No Liveness component ePassport at SmartGate Kiosk	Liveness component ePassport
V2	2	Detection component SmartGate Ticket issued by Kiosk	Detection component ePassport
V3	3	Token Read	Token Read
	4		
	5		
V4	6	Facial Verification	Facial Verification
	7		
	8		
V5	9	Local	Local

In 2010, a report from Frontex Europe [8] noted that over one million travellers had used SmartGate at all airports since opening in 2007. Of the 200 travellers who were interviewed in the report, 98% agreed that the process was easy, 97% agreed that they were extremely likely to recommend SmartGate and 96% agreed they would use SmartGate again. Typical causes of false rejections include users not looking directly the camera or poor quality photos stored in electronic passports. During peak traffic, passenger interactions tended to improve with individuals learning from other users and then repeating the learnt behaviours.

In 2013, the next generation of the Smart Gate system, SmartGate+, was trialled at Auckland Airport in New Zealand. The configuration changed from two to a one-step configuration which removed the kiosk component and matched a configuration similar to the standard EU e-Gate system (an arrangement that accounts for 43% of global deployments) [35]). Table 4 describes the major verification differences between the SmartGate and SmartGate+ configurations using the GM and ICP.

In an 18 hour period of SmartGate+ operation, 400 unique users made a total of 449 separate interactions (some users re-entered the system after a rejection - subsequent attempts were counted as separate interactions). All travellers using the system were arrivals who were not pre-selected beforehand. The only requirement was that travellers using the system were over the age of 18 and were either Australian or New Zealand citizens with a biometric passport. Airport personnel were on-hand to provide assistance before and during the ABC system use.

Traveller interactions were observed at a distance using CCTV cameras. The instructions displayed on the monitor were clear from a front view of the gate. Noticeable user behaviours are defined when the user makes a movement or gesture that can be clearly seen and should have an impact on their performance.

A total of 367 users (81.73%) were accepted through the system. The remaining 82 users (18.26%) had to leave the ABC implementation and queue for manual inspection. 49 of the rejected users (10.91%) were observed not to contribute any noticeable user errors during their attempt. We can, therefore, attribute this to either ineligibility or system error/maintenance. A total of 33 users (7.35%) were correctly rejected for making clear erroneous presentations such as not facing the camera during facial acquisition (24 users) or taking the passport out before reading had finished (6 users). However, the system can compensate for minor deviations in performance. 106 subjects who were accepted made noticeable user errors such as face movements (64) and re-entering the passport during reading (32).

Of the 449 interactions, 66 required some form of assistance from personnel, while 78 travellers failed to step incorrectly on the foot signs (marks on the floor of the gate illustrating where to place feet while looking at the camera). Of those 78 subjects, 54 were accepted for facial verification which meant the system was able to identify an acceptable biometric sample. The remaining 24 were not in range of the camera which resulted in Defective Interactions.

Conventionally the SmartGate+ system would report an FRR of 18.26%. However, through observation, we have concluded that 10.91% included system error while user interaction errors caused 7.35% of the rejects. Implementing the full HBSI Model would allow a further breakdown of these performance measures, detailing specific categorisations in system processing which we were unable to determine during the trial. It is also important to note that it was assumed that every user was genuine for this trial, and hence the FAR metric was not calculated.

TABLE V
CATEGORISING USER BEHAVIOUR WITH ICP AND GM

GM	ICP	Desired Behaviours	Cautious Behaviours	Undesired Behaviours
V1	1	Enter booth promptly (424) Adjust to stand on foot signs, pose correctly and follow instructions (414) Locate passport (425)	Enter booth too quickly after previous traveller (25) Juggling passports, tickets bags (19)	Block gate with luggage/leave luggage outside of gate (8) Try to enter through closed gate (4) No search for Passport (15)
V2	2	Traveller makes a correct presentation (e.g. Machine Readable Zone downwards on the reader, passport cover off). Checking token that it is fully inserted (402)	Moving passport within reader (16)	Enter the wrong token, entered plane ticket, entered token upside down/incorrectly – not aware of this (7)
V3	3	Traveller patiently watches/listens for the next instructions (415)	Moves hand to token reader in anticipation (12)	Takes passport out before reading is completed or before instructed (24)
V3	4	Internal System Process – behaviours the same as Step 3		
V3	5	Internal System Process – behaviours the same as Step 3		
V4	6	User understands where the camera is and is aware of where to look User faces camera and keeps head still for a system determined amount of time (419)	User is searching for the camera or does not understand where the verification is taking place (49) User is distracted throughout interaction and loses focus on looking at the camera (32)	The user is distracted for too long and system times out. (18) The user is unaware of process and verification fails. (12) The user is continually moving throughout the process. (26)
V4	7	Internal System Process – behaviours the same as Step 6		
V4	8	Internal System Process – behaviours the same as Step 6		
V5	9	User exits booth when prompted (425)	User waits for more information (6) User spends time putting away passport, sorting luggage (25)	User does not move (6)

4.2 Categorising User Behaviour

Observing a total of 449 interactions, we noted traveller behaviour throughout the SmartGate+ trial. In Table 5 we harness the ICP and GM to establish desirable, cautious and undesirable behaviours for each step of the process. The table highlights observed behaviours for a particular step, stating the number of instances a behaviour was witnessed during a particular task.

While many users performed desirable behaviours, there were a relatively high number of 'bad' behaviours performed. There were clear instances of users who made no action or showed no knowledge of the system process (for example, some users try to enter through closed gates or were unaware that they had to present their passport in Steps 1-3).

Categorising potential behaviours and noting the number of occurrences throughout a trial will help to identify the system's ability to handle that specific behaviour and in the future, allow options for feedback. For example, if a system was able to classify that 26 users are continually moving throughout the facial interaction, then appropriate feedback could be displayed in an attempt to correct the presentation. Likewise, if a user was smiling or were wearing glasses, an image processing element could relay the information to stop smiling and/or remove their glasses to comply with ISO standards.

4.3 Applying the HBSI Presentation Framework

By using the GM and ICP, we can categorise all potential user behaviours and possible system handlings within a particular scenario. Once we have determined the possible outcomes for a particular behaviour, it is feasible to apply the relevant HBSI Presentation Framework categorisations. Table 6 presents some scenarios where user behaviour has been defined by system handling and the corresponding HBSI metrics. In the first scenario, for example, an unwanted behaviour of the user-facing away from the camera and not being aware of this interaction occurred twelve times throughout the SmartGate+ trial. We observed these scenarios when the information on the monitor was requesting the user to look at the camera. The system can handle this situation in two possible ways. The first instance is where the system can identify correctly that the biometric was not presented and therefore by applying the HBSI framework; classify the presentation as an FI. If the system incorrectly determines the user has submitted their biometric, a system process should flag the sample as unsuitable for the next stage. HBSI indicates a DI (as this was a user error).

The development of this model would allow for real-time feedback. For example, in this scenario, a DI categorisation could alert and train the user how to perform the correct behaviour, therefore reducing the likelihood of the error in the future. If an FI were categorised, then an appropriate response from the system can be made. Feedback should assist the user in correcting their presentation.

In this scenario, the users were correctly rejected by the system for not providing a suitable biometric sample. Nine travellers were observed to either repeat the entire process (as they were completely unaware of the facial verification process) and three withdrew from the ABC system in favour for manual control. For False Interactions, we were able to determine seven instances where the user took out their passport before reading, eight who faced away from facial verification and five for various other reasons, but were all successfully detected by the system as erroneous.

TABLE VI
AN EXAMPLE CATEGORISING USER BEHAVIOUR WITH THE HBSI FRAMEWORK

GM	ICP	Behaviour	Possible System Handling	HBSI	Notes
V4	6	Unaware of process, facing away from camera (12)	Biometric Presented	Defective Interaction	It is evident in this example that the user made an incorrect presentation.
			Biometric Not Presented	False Interaction	
V4	6	User understands where the camera is and is aware of where to look (409)	Biometric Presented	Successfully Processed Sample.	In this example, it is evident the user makes a correct presentation. The ideal outcome is a Successfully Processed Sample however if there is an FTA or FTP error the system will determine this.
		User faces camera and keeps head still for a system determined amount of time (419)	Biometric Not Presented	Failure To Acquire or Failure To Process	
			Biometric Presented	Successfully Processed Sample, FTA or FTP	
V4	6	User is distracted throughout interaction and loses focus on looking at the camera (32)	Biometric Not Presented	Defective Interaction, Concealed Interaction or False Interaction	Cautious behaviours are difficult to classify as a correct or incorrect behaviour. The outcome of this action will largely be dependent on the capture time of the camera and depending on the sample taken, wherever it meets the templates requirements.

A total of 367 travellers (81.73%) were accepted through the gate. 213 travellers (58.13%) were accepted with little or no issues (and, therefore, were categorised as Successfully Processed Samples). We recorded 86 (23.61%) users were accepted but made noticeable interaction errors. Although some of these mistakes were minor in the sense that the sensor should be able to account for a particular behaviour, we estimate that some these interactions should have been classified as Concealed Interactions. Over the course of the trial, the system accepted noticeable erroneous presentations on six occasions. In these instances, travellers were looking away from the camera throughout the facial capture. The researchers observed that traveller's faces were yaw-rotated in such a way from the camera that successful capture was unlikely. However, the system accepted the sample and allowed for successful border passing.

Ideally, a Successfully Process Sample (SPS) is expected when a user performs a sequence of desired behaviours for all token and biometric presentations to a sensor. The ultimate goal for an ABC system, therefore, is for all users to perform optimally. However, due to whatever reason, if cautious or undesired behaviours are made then they can be addressed at each step of the GM and ICP. Appropriate categorisations can then be made and steps can be taken to resolve these issues in future implementations.

4.4 Combining Usability Assessment with the HBSI Evaluation Method

During this data collection, all travellers were offered the chance to participate in a questionnaire after passing through the gate. 35 subjects participated providing comments which were used to measure user satisfaction.

Effectiveness was measured using the total number of errors made which was 139 (30.96%) out of the 449 interactions observed. The number of assists was at 66 (14.92%) and 367 users (81.73%) completed the task of the border crossing. Effiency in terms of task time was measured on average at 17 seconds. Users who made no errors during their task completed on average in 10 seconds whilst travellers who made incorrect presentations saw task times extended to on average 26 seconds.

While reporting usability performance provides a powerful tool for measuring the quality of the user experience, these tests are often performed in a controlled environment whereby a researcher is present. Components of usability evaluations have several disadvantages; for example, during implementation, it is not feasible for user satisfaction to be captured on a consistent basis. Usability assessments can be fairly complex and time-consuming to analyse. Travellers may also be tired, stressed or in a rush to check through the border and therefore not willing to participate in any open-ended questionnaires.

It is imperative that errors, user or system generated, are classified and analysed throughout the entire systems lifecycle. Assessing incorrect interaction from a system perspective will aid in highlighting where potential problems can occur. Our method synergies with other usability evaluations. For example, through usability assessment, we measure that some travellers are having problems with passport reading. In our proposed approach the framework will attempt to explain why this may be occurring. It could be due to a sensor or processing fault or perhaps because of an incorrect user interaction. Ultimately this

will enable a deeper analysis of the performance. A combination of HBSI and usability performance assessment will only advance future design and implementation of these systems.

The models proposed in this study outline clear methodologies to categorise incorrect user interactions and system errors. Defining potential cautious or unwanted behaviours using tracking hardware such as the Microsoft Kinect or image processing elements in these complex systems will allow future implementations to adapt to the user, improving the user experience and, therefore, reducing error rates.

V. Conclusions

Border controls across the globe are progressively installing ABC systems to improve security, streamline the travelling process and working towards facilitating a better passenger experience. In this paper, we have identified system and user interaction problems and where processing faults may lie.

We have proposed a Generic Model, which can be used to standardise the mapping of ABC configurations to identify where variations and similarities lie between configurations. Having defined a system, we can investigate individual interactions within the verification process by applying the Identity Claim Process. Allowing the study of each step of the verification process to identify conditions where the HBSI Presentation Framework can be implemented. Identifying common scenarios and noting how a particular system handles certain behaviours will be useful in highlighting bottlenecks within a process. The HBSI evaluation method also considers usability and ergonomic variables that also attribute to system performance.

Standard measures such as FTA, FTE and FRR, can sometimes mask the true reason behind why an error occurred. Harnessing the HBSI method allows for a full categorisation of a range of metrics which will benefit in analysing system performance.

We have measured usability metrics from the SmartGate+ trial through analysing user satisfaction, effectiveness and efficiency to understand how the system performs at a human-biometric level. 139 interactions (30.96%) of the 449 included some major user interaction errors such as travellers not knowing what to do, facing away from the camera or taking out their passport before the read process was complete. Reasons behind these behaviours could be tiredness, stress or that travellers simply do not understand the process. Out of the 139 interactions, 33 were correctly rejected from the system for making an incorrect presentation to the system. 86 interactions were accepted when undesired/cautious behaviours were presented. Therefore 20 interactions contained minor errors but were still accepted by the system.

Evaluating user behaviour for each task and mapping out all possible scenarios within the system will be crucial to configuring and adapting system responses. By applying the HBSI framework, this may lead to an enhanced system performance, helping to reduce errors, and enhance overall usability of ABC systems for travellers worldwide.

Implementing our proposed work will enable automated assessment of traveller interactions. Introducing methods to assess the user through the introduction of new tools such as eye tracking or image processing elements will provide many benefits. For example, for the tired or stressed traveller, introducing an automated feedback system to relay information to the user on how to correct their presentation, e.g. look up, open eyes, will begin the first steps into offsetting incorrect behaviours. Further work will be needed to identify common presentations and appropriate methods in responding to users making incorrect interactions. Current HBSI work is investigating the use of skeletal tracking and image processing in an attempt to improve user presentation. Also, research is studying the different processes of communicating this feedback back to the user (e.g. through text or icons)

The proposed research method systematically decomposes an ABC system and identifies the process flow. The models proposed in this paper build upon live data captured from the trial of the next generation of border control applications, simulating ideal experimental conditions. The results of our assessment thus have external validity. The implementation of the well-established HBSI framework builds upon on successful testing of single biometric modalities. This paper has taken the first steps into complex multi-modal reporting. An advantage HBSI has over others performance assessment methods does not only does it allow a deeper understanding of a reason behind a failure, but with the right technological advancements, can be used throughout the products life cycle.

Further work will be required, however, to attribute HBSI to the use of token presentations and other processes such as a user entering or exiting the system. Additionally, further data will be required from a wider range of participants and other live implementations to validate our approach. More work is needed in categorising user behaviours and the effects these have on the system. What constitutes as a ‘noticeable user error’ must also be considered, especially within an image processing environment. Care must also be taken in operational testing to make sure that the live scenario of border control is replicated as closely as possible. In a controlled environment, influencing factors such as stress or tiredness will not be able to be replicated.

In conclusion, this research may contribute to improvement in the accuracy of reporting of system performance in ABC systems. The application of the HBSI framework will allow a range of metrics, defining a set of interaction measurements which must be a priority (while adhering to the systems intended use) in the design and implementation of these public systems. Reporting on the six HBSI presentation metrics will allow a deeper understanding of where problems lie within a system. The models proposed will enable the breakdown of the process so that each stage can be assessed beyond the traditional reporting of a system level error. In defining a process map, user and system handlings are measured at each key component.

[1] M.Nuppeney, “Automated Border Control - State of Play and Latest Developments,” in *NIST IBPC* 2014. [Online]. Available: http://www.nist.gob/itl/iad/ig/ibpc2014.cfm.

[2] Frontex Europa, “Best Practice Operational Guidelines for Automated Border Control (ABC) systems,” 31 August 2012. [Online]. Available: http://frontex.europa.eu/assets/Publications/Research/Best_Practice_Operational_Guidelines_for_Automated_Border_Control.pdf.

[3] Department of Transport, UK GOV, “Public Experiences of and Attitudes towards Air Travel,” 29 July 2010. [Online]. Available: https://www.gov.uk/government/publications/public-experiences-of-and-attitudes-towards-air-travel.

[4] International Organization for Standardization (ISO), "Information Processing System - Vocabulary - Part 37: Harmonized Biometric Vocabulary," vol. DIS, no. 2382.37, 2010.

[5] Frontex Europa, "BIOPASS Study on Automated Biometric Border Crossing Systems for Registered Passengers," European Agency for the Management of Operational and Cooperation at the External Borders, Warsaw, 2007.

[6] M.Drahansky, "Liveness Detection in Biometrics," in *Advanced Biometric Technologies*, Brno, Czech Republic , InTech, 2011, pp. 439-444.

[7] ISO/IEC, "92411-11 Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs) - Part 11 Guidance on Usability," ISO/IEC 9241-11, 1998.

[8] Frontex Europa, "BIOPASS II Automated Biometric Border Crossing Systems Based on electronic passports and facial recognition," Frontex, Warsaw, 2010.

[9] M. Theofanos, R. J. Michaels and B. C. Stanton, "Biometric Systems Include Users." *IEEE Systems Journal*, vol. 3, no. 4, pp. 461-468, 2009.

[10] M. Theofanos, B. Stanton, R. Michaels and S. Orandi, "Biometric Systematic Uncertainty and the User," in *Proc. First IEEE International Conference on Biometrics: Theory, Applications, and Systems*, Crystal City, VA, 2007.

[11] R. J. Michaels, B. Stanton, M. Theofanos and S. Orandi, "A taxonomy of Definitions for Usability Studies in Biometrics," US Department of Commerce, National Institute of Standards and Technology, 2006.

[12] Y. Y. Choong, M. Theofanos and H. Guan, "Fingerprint Self-Captures: Usability of a fingerprint system with real-time feedback," in *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems*, 2012.

[13] B. Fernandez-Saavedra, R. Sanchez-Reillo, J. Liu-Jimenez and O. Miguel-Hurtado, "Evaluation of biometric system performance in the context of Common Criteria," *Information Sciences*, vol. 245, no. 1, pp. 240-254, 2013.

[14] B. Fernandez-Saavedra, R. Alonso-Moreno, A. Mendaza-Ormaza and R. Sanchez-Reillo, "Usability Evaluation of Fingerprint Based Access Control Systems," in *Proc. Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP*), 2010 Sixth International Conference on, Darmstadt, 2010.

[15] M. Ylikauppila, S. Toivonen, M. Kulju and M. Jokela, "Understanding the Factors Affecting UX and Technology Acceptance in the Context of Automated Border Controls," in *Proc. IEEE Joint Intelligence and Security Informatics Conference (JISIC*) 2014.

[16] B. Fernandez-Saavedra, R. Alonso-Morendo, J. Uriart-Antoniou and R. Sanchez-Reillo, "Evaluation methodology for analyzing usability factors in biometrics," in *Proc. Security Technology,. 43rd Annual 2009 International Carnahan Conference*, 2009.

[17] A. Sasse, "Red-Eye Blink, Bendy Shuffle and the Yuck Factor. A User Experience of Biometric Airport Systems," *IEEE Security and Privacy*, vol. 5, no. 3, pp. 78-81, 2007.

[18] A. Adhmad and A. Ali, "Iris Recognition and the Challenge of Homeland and Border Control Security in UAE," *Telematics and Informatics*, vol. 25, no. 2, pp. 117-132, 2008.

19] Öffentlicher Abschlussbericht, "Untersuchung der Leistungsfähigkeit von biometrischen Verifikationssystemen – BioP II (A Study of the Peformance of Biometric Systems)," *Secunet, Essen*, 2005.

[20] UKPS, "UK Passport Service Biometrics Enrolment Trial," Atos Origin, London, 2005.

[21] Air India, "Automated Passport Control Process Guide," 30 July 2014. [Online]. Available: http://airindia.in/Images/pdf/Redsigned-Automated-Passport-Control.pdf.

[22] B. Shackel, "Usability Context, Framework, Definition, Design and Evaluation," *Human Factors for Informatics Usability*, vol. 21, no. 5-6, pp. 339-346, 2009.

[23] M. El-Abed, R. Giot, B. Hemery and C. Rosenberger, "A Study of User's Acceptance and Satisfaction of Biometric Systems," in *Proc. Security Technology (ICCST) IEEE International Carnahan Conference*, San Jose, CA, 2010.

[24] A. Kriechbaum and M. Clabian, "FBC and FastPass - Future Border Control, A Harmonized, Modular Reference System for all European Automatic Border Crossing Points," in *KIRAS 2013*, Austria, 2013.

[25] European Commission, "ABC4EU Project Overview," June 2014. [Online]. Available: http://abc4eu.com/about/.

[26] K. Singla and S. Kumar, "A Review of Data Acquisition and Difficulties in Sensor Module of Biometric Systems," Songklankarin *Journal of Science and Technology*, vol. 5, no. 35, pp. 589-597, 2013.

[27] D. Cantaero, D. Herrero and F. Mendez, "A Multi-Modal Biometric Fusion Implementation for ABC Systems," in *Proc. European Intelligence and Security Informatics* Conference, Alcobendas, Spain, 2013.

[28] E. Kukula, S. Elliott and V. Duffy, "The Effects of Human Interaction on Biometric System Performance," in *Digital Human Modelling*, Springer, 2007, pp. 904-914.

[29] E. Kukula, "Understanding the Impact of Human-Biometric Sensor Interaction and System Design on Biometric Image Quality," in *NIST Biometric Quality Workshop II*, Purdue, 2007.

[30] E. Kukula and S. Elliott, "A Definitional Framework for the Human-Biometric Sensor Interaction Model," in *Proc. SPIE7667, Biometric Technology for Human Identification VII*, 2010.

[31] E. Kukula, "Framework for Human, System, and Administrative Errors in Biometric Systems," in *INCITS*, Washington, DC, 2009.

[32] S. Elliott, K. O'Connor, J. Robertson and R. Guest, "Expanding The Human-Biometric Sensor Interaction Model to Identity Claim Scenarios," *in Proc. IEEE International Conference on Identity, Security and Behaviour Analysis (ISBA)* Hong Kong, 2015.

[33] NIST, "Usability & Biometrics: Ensuring Successful Biometric Systems," National Institute of Standards and Technology, 2008.

[34] ONE News, "SmartGate open to US and UK travellers," 29 July 2013. [Online]. Available: http://tvnz.co.nz/travel-news/smartgate-open-us-and-uk-travellers-5521551.

[35] Acuity Market Intelligence, "Global Airport ABC eGates & Kiosks Revenue," 22nd April 2014. [Online]. Available: http://www.prweb.com/releases/2014/03/prweb11635040.htm. [Accessed 8 August 2014].

[36] D. O. Gorodnichy, S. Eastwood, V. Shmerko and S. Yanushkevich, "Automated Border Control Systems as Part of e-Border Crossing Process," in National Insistute of Standards - International Biometric Performance Conference (IBPC 2014), Gaithersburg, 2014.

[37] ISO/IEC, "ISO/IEC FCD 19795-6, Information Technology - Biometric Performance Testing and Reporting - Part 6:"

Joshua J. Robertson received his BEng degree in Computer Systems Engineering from The University of Kent, Canterbury, United Kingdom in 2013. He is currently a graduate researcher at the University of Kent studying a PhD in Image and Information Engineering. His research interests involve biometrics, border control and human errors. His work investigates the area of usability when applied to Automated Border Control Systems.

Richard M. Guest obtained his PhD in 2000 and has been a member of academic staff (currently Senior Lecturer) at the University of Kent since this date. He is Deputy Head of the School of Engineering and Digital Arts. His research interests lie broadly within image processing and pattern recognition, specializing in biometric and forensic systems, particularly in the areas of image and behavioural information analysis, standardization and document processing.

Stephen J. Elliott is an associate professor with an appointment in Technology, Leadership and Innovation at Purdue University, where he has been a member of the faculty since 2001. He is the Director of the International Center for Biometric Research (ICBR).

Kevin O'Connor is the managing director of the International Center for Biometric Research (ICBR) at Purdue University. His research has encompassed a number of different modalities, including face and fingerprint recognition, and he is credited with developing the Stability Score Index for biometric performance.