# A study of users' acceptance and satisfaction of biometric systems

Mohamad El-Abed, Romain Giot, Baptiste Hemery, Christophe Rosenberger

# A study of users' acceptance and satisfaction of biometric systems

Mohamad El-Abed, Romain Giot, Baptiste Hemery and Christophe Rosenberger
GREYC Laboratory, ENSICAEN - University of Caen Basse-Normandie - CNRS
ENSICAEN, 6 boulevard marechal Juin 14050 Caen Cedex 4 FRANCE
Email: {mohamad.elabed, romain.giot, baptiste.hemery, christophe.rosenberger}@greyc.ensicaen.fr

***Abstract*—Biometric authentication methods are being increasingly used for many types of applications. Since such methods necessitate humans to interact with a device, effective implementation requires consideration of the perceptions and responses of end users. Towards this goal, we present in this paper a modality-independent evaluation methodology to study users' acceptance and satisfaction of biometric systems. It is based on the use of a questionnaire and some data mining tools for the analysis. We have applied it on two biometric systems developed in our research laboratory. The results from this study demonstrated that users' satisfaction analysis should be more taken into account when developing biometric systems. A significant panel of 70 users was more satisfied from the keystroke system than the other one. Users surprisingly considered that its perceived performance was also better even if the used face system has a better performance with an EER of 8.76% than the keystroke one with an EER of 17.51%. The robustness of a system against attacks and its perceived trust have been identified as important factors to take into account when designing biometric systems. Results have also demonstrated significant relationships between demographic factors and their perception about the biometric technology and the studied systems.**

***Index Terms*—evaluation of biometric systems, users' acceptance, Kruskall-Wallis test, Bayesian networks, Decision Trees.**

## I. Introduction

Biometrics offers automated methods for identity verification and identification based on physiological, morphological or behavioral characteristics. By contrast with possession-based ("what we own" such as a key) or knowledge-based ("what we know" such as a password), biometrics is based on "what we are" and "how we behave". Many biometric systems have been proposed in the literature for the last decade [1], [2]. They are mainly used to manage the access of physical (e.g., airports) and logical (e.g., e-commerce) resources. Evaluating biometric systems constitutes one of the main challenges in this research field. Nowadays, many studies have been done to evaluate such systems. Evaluation is generally realized within four aspects as illustrated in Fig. 1:

- *performance*: defines some quantitative metrics to measure the efficiency of biometric systems [3] such as equal error rate (EER), failure-to-enroll rate (FTE), computation time ...;
- *acceptability and user satisfaction*: measures users' perception, feelings and opinions regarding the system [4];
- *data quality*: measures the quality of the biometric raw data. Low quality samples increase the enrolment failure rate, and decrease the system performance;
- *security*: measures how well a biometric system (*algorithms* and *devices*) can resist to several types of attacks. Different attacks have been reported and analyzed in [5] such as presenting a fake (e.g., dummy finger), replay attack, Denial of Service (DoS) ...

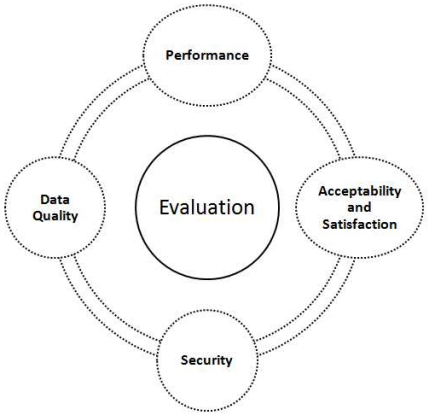


Fig. 1. Evaluation aspects of biometric systems: performance, acceptability and satisfaction, data quality and security.

The evaluation of biometric systems is now carefully considered. Many databases have been collected (such as UBIRIS [6]), many competitions (such as Fingerprint Verification Competition [7]) and platforms have been proposed (such as BioSecure [8]) whose objective is mainly to compare enrollment and verification/identification algorithms in the state of the art. Multiple metrics are used within this context [3]. These statistical measures allow in general a precise performance characterization of a biometric system. Nevertheless, these works are dedicated to quantify performance technology (algorithms, processing time, memory required ...) without taking into account users' perception when using these systems. One important challenge is that a biometric system is human centric [9]. The authors in [10] illustrate the complexity of designing a biometric system on three main factors: (i) accuracy in terms of errors, (ii) scale or size of the database and (iii) usability in terms ease of use, acceptability ... Therefore, studying and

analyzing users' perception constitutes a key factor to make it operational and acceptable. One government can decide that an individual would be identified through a biometric data embedded in the passport. For logical or physical access control in a company, it is more difficult to impose a system that would be not accepted by users. As for example, DNA analysis is one of the most efficient techniques to verify the identity of an individual or to identify him. Nevertheless, it cannot be used for logical or physical access control not only for time computation reasons, but also because nobody would be ready to give some blood to make the verification.

Nowadays, there is a lack of a generic evaluation methodology that takes into account users' perception within the evaluation process, which constitutes one of the main drawbacks for biometric systems proliferation. For this purpose, we propose in this paper a *modality-independent evaluation methodology* to study users' acceptance and satisfaction of biometric systems. Such kind of evaluation, in our opinion, will: (i) enhance the performance of biometric systems [11], (ii) improve the accuracy of the optimistic results provided by biometric system designers in terms of errors (e.g., EER) and (iii) reduce product complexity and increase user satisfaction.

The paper is organized as follows. In section II, we present the background of several works done to understand the different factors affecting users' acceptability and satisfaction. Section III details the proposed methodology based on user survey. Experimental results of the proposed methodology are given in section IV. We give a conclusion and propose some perspectives of this work in section V.

## II. Background

Understanding human perception can be exploited in different domains. Human perception is quite predictable in many instances and can be understood with a high degree of accuracy. Extracting such information aims to add the community in biometrics to: (i) more understand the needs of users and (ii) improve the quality of biometric systems (*algorithms* and *devices*). Nowadays, several studies have been done to quantify users' acceptability and satisfaction of biometric systems such as:

- NIST Biometrics Usability group has performed a usability test on fingerprints [12]. The survey was conducted on 300 adults recruited from a pool of 10,000 people. There were 151 women and 149 men ranging in ages from 18 to over 65 years. 77% of participants were in favor to provide fingerprint images as a mean of establishing identity for passport purposes. 2% of participants have expressed concerns about the cleanliness of the devices with which they would have physical contact. Another study has been done by NIST to examine the impact on fingerprint capture performance of angling the fingerprint scanners (flat, 10, 20 and 30 degrees) on the existing counter heights (99, 114.3 and 124.5 cm) is presented in [13];
- Opinion Research Corporation International (ORC International) has presented in [14] the results of a phone survey conducted on 2001 and 2002. The survey has been conducted among national probability samples of 1017 and 1046 adults, respectively, living in United States. The 2001 study showed that 77% of individuals feel that finger-imaging protects individuals against fraud. For privacy issues, 87% in 2001 and 88% in 2002 are worried for the misuse of personal information. There is a good percentage of acceptance, more than 75%, for U.S. law enforcement authorities requiring fingerprint scans to verify identity for passports, at airport check-ins and to obtain a driver license (see [14] for more details).
- Other studies presented in [15], [16], [17], [18], [19], [20], [21], [22], [23], [24] have highlighted several points about biometrics such as:
  - acceptance is linked to the number of uses of the biometrics in general and information provided by the biometric device can also improve user acceptance;
  - there is a potential concern about the misuse of personal data (i.e., templates) which is seen as violating users' privacy and civil liberties. Another important concern is the probability that criminals may perpetrate heinous acts to gain access. This could include stalking or assaulting individuals to steal their biometric information;
  - individuals complain that once the biometric template is stolen, it is compromised forever;
  - there are also concerns about hygiene with touching such devices and health risks for more advanced technologies such as iris or retina. According to our knowledge, none paper has emphasized physical harm to users of these systems. But despite of this, several concerns were highlighted along this interaction. Anecdotally, some users of biometrics have complained that hand geometry systems dry their hands while military aviators participating in an experimental program voiced concern that retinal scanning would damage their vision with extended use over time.

### A. Discussion

Studies presented in the previous section highlighted different important factors that have to be taken into account when studying users' perception such as:

- *socio-demographic factors:* such as age, gender, ethnicity, religion, experience and ability;
- *learnability and memorability:* they mainly concern how rapidly a user can use the system after instruction or training;
- *confidence or trust:* indicates how the performance of the system is perceived by users. It depends mainly on feedbacks from users and their experience;
- *ease to use:* depends on the quality of the biometric sensor and the ergonomic interface. It may also depend on the time required for verification or identification. For example, if the biometric system takes several minutes

between the acquisition of the required data and user identification, users may believe that the biometric system is not easy to use;
- *Privacy issues:* there is a potential risk concerning the misuse of the personal collected data, which is seen as violating user's privacy and civil liberties. Many debates have been conducted over the central storage of biometric templates versus holding the personal template on a smart card where the verification is locally processed;
- *Physical invasiveness:* the acquisition of biometric data requires user interaction with the biometric sensor. Depending on the used method, acquisition of biometric data is performed with or without contact with the biometric sensor;
- *Cultural issues:* the acceptability denotes the way how users perceive the biometric system and interact with it. Acceptability is highly dependent of the culture of users. As for example, cultures with an aversion to touch public surfaces would prefer to use biometric systems with contactless sensors (e.g., iris or palm veins).

Studies done on evaluating users' acceptability and satisfaction of biometric systems are very few in comparison with performance ones. Moreover, these studies are based on statistical answers to a questionnaire but no serious data analysis is conducted for understanding the reasons. In order to contribute to solve this problem, the goal of this paper is to propose a methodology that studies users' perception to enhance the usability of biometric systems. This evaluation methodology is modality-independent (i.e., it could be applied on any kind of biometric systems). As for us, taking into account users' view of the biometric process (hardware, software and instructional design) is not only beneficial to the end users but it will also help to improve the performance and effectiveness of a system [11].

## III. Evaluation Methodology

The proposed methodology was designed to quantify users' satisfaction on the use of biometric systems. To accomplish this objective, we developed a survey instrument for data collection. These kinds of surveys enable to gather information to be statistically analyzed. The proposed methodology principle is illustrated in Fig. 2: It collects the data using a questionnaire (section III-A). This step is followed by a pre-processing phase to extract the significant knowledge (section III-B). Then, the Kruskall-Wallis test is performed to determine if there is a significant relationship between demographic characteristics and respondents' answers (section III-C). Data mining tools are used to explain these answers to determine the possible candidates that affect their acceptance and satisfaction of biometric systems (section III-D).

### A. Data collection

The data collection phase is based on a survey instrument (see Appendix). It was designed to collect demographic, experiential and attitudinal characteristics that might have an impact on or a relationship to respondents' views on the use of biometrics. The survey takes into account other works in the field done by NIST [13], NCR [25], ORC [14], Atos origin [26] and other works presented in [27], [20], [19], [21]. It also noted unsolicited questions that we found it valuable when collecting the GREYC-Keystroke database [28]. A 4-point Likert-type scale is used to valuate the answers on the perception questions. The survey questionnaire contains 18 questions divided into two sets:

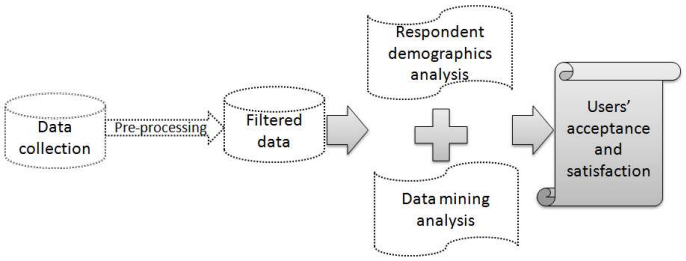


Fig. 2. Methodology principle

- **general perception of biometric systems** (Appendix, part B) which contains 7 questions aiming to understand users' experience on biometric technology;
- **perception of the tested system** (Appendix, part C) which contains 11 questions aiming to measure users' satisfaction and acceptance on the tested system.

In addition to these questions, we request some information on the individual: gender, age ...(Appendix, part A). These demographic characteristics are requested to determine if there is significant relationships between them and respondents' perception on biometric technology. Also, we propose the question 16 (Appendix, part C) to identify where the use of the tested system would be appropriate.

### B. Data Pre-processing

Prior to analyzing the pilot data, we use a technique to enhance the accuracy and the reliability of the extracted knowledge. It consists of deleting answers having a predefined number of gaps (i.e., questions without any answers).

### C. Respondent demographics analysis

In order to determine whether there is a significant relationship between demographic characteristics and respondents' perception on biometric technology, we use the Kruskall-Wallis test (KW). It is a nonparametric (distribution free) test, which is used to decide whether K independent samples are from the same population. In other words, it is used to test two hypothesis given by Eq. 1: the null hypothesis $H_0$ assumes that samples originate from the same population (i.e., equal population means) against the alternative hypothesis $H_1$ which assumes that there is a statistically significant difference between at least two of the subgroups.

$$\begin{cases} H0: \ \mu_1 \ = \ \mu_2 \ = ... = \ \mu k \\ H1: \ \mu_i \ \neq \ \mu_j \end{cases} \quad (1)$$

The Kruskal-Wallis test statistic is given by Eq. 2 and the p-value is approximated, using chi-square probability distribution, by $Pr(\chi^2_{g-1} \geq H)$. The decision criterion to choose the appropriate hypothesis is defined in Eq. 3.

$$H \ = \ \frac{12}{N(N+1)} \sum_{i=1}^{g} n_i \ \bar{r_{i.}^2} \ - \ 3 \ (N+1) \quad (2)$$

where $n_i$ is the number of observations in group $i$, $r_{ij}$ is the rank of observation $j$ from group $i$, $N$ is the total number of observations across all groups.

$$\bar{r_{i.}^2} \ = \ \frac{\sum_{j=1}^{n_i} r_{ij}}{ni} \ and \ \bar{r} \ = \ \frac{1}{2} \ (N+1)$$

$$\begin{cases} p-value \ \geq \ 0.05 & accept \ H_0 \\ otherwise & reject \ H_0 \end{cases} \quad (3)$$

*D. Data mining analysis*

In order to analyze respondents' answers, we use two types of classifiers: *Bayesian networks* [29] and *Decision Trees* [30]. They are formal graphical tools for representation of decision scenarios requiring reasoning under uncertainty.

*1) Bayesian networks:* A Bayesian network $(B_S, B_P)$ is a probabilistic graphical model that represents a set of random variables $U = \{x_1, x_2, ...x_n\}$ and their conditional independencies. The Bayesian structure $B_S$, is a directed acyclic graph (DAG) where nodes represent propositional variables in a domain, and the arcs between nodes represent the dependency relationships among the variables. The Bayesian probability distributions $B_P$, is a set of probability tables $B_P = \{p(u|pa(u)) \ | u \in U\}$ where $pa(u)$ is the set of parents of $u$ in $B_S$.

The method used to learn the bayesian network structure $B_S$ is based on *conditional independence tests* as described in [31]. This method mainly stem from the goal of uncovering causal structure. The assumption is that there is a network structure that exactly represents the independencies in the distribution that generated the data. The method is divided into two stages:

- **find a skeleton:** starting with a complete undirected graph, the method try to find conditional independencies $\{x \rightarrow y\} \cup \forall z \in Z \ z \rightarrow y$ in the data. If a, independency is identified, the edge between $x$ and $y$ is removed from the skeleton. To test whether variables x and y are conditionally independent given a set of variables Z, a network structure with arrows $\forall z \in Z \ z \rightarrow y$ is compared with one with arrows $\{x \rightarrow y\} \cup \forall z \in Z \ z \rightarrow y$. A test is performed by using a predefined score metric. In this study, we use four score metrics as defined below. We use the following conventions to identify counts in the database $D$ of a network structure $B_S$. Let $r_i$ $(1 \leq i \leq n)$ be the cardinality of the variables $x_i$. We denote by $q_i$ the cardinality of the parent set of $x_i$ in the network structure $B_S$. Hence, $q_i$ can be calculated as the product of cardinalities of nodes in $pa(x_i)$, $q_i = \prod_{x_j \in pa(x_i)} r_j$. We denote by $N_{ij}$ $(1 \leq i \leq n, \ 1 \leq j \leq q_i)$ the number of records in $D$ for which $pa(x_i)$ takes its $j$th value. We denote by $N_{ijk}$ $(1 \leq i \leq n, \ 1 \leq j \leq q_i, \ 1 \leq k \leq r_i)$ the number of records in $D$ for which $pa(x_i)$ takes its $j$th value and for which $x_i$ takes its $k$th value. Hence, $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$. We use $N$ to denote the number of records in $D$.
  - Entropy metric $H(B_S, D)$ defined as

$$H(B_S, D) = -N \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \frac{N_{ijk}}{N} \log \frac{N_{ijk}}{N_{ij}} \quad (4)$$

  - AIC metric $Q_{AIC}(B_S, D)$ defined as

$$Q_{AIC}(B_S, D) = H(B_S, D) + K \quad (5)$$

    where $K = \sum_{i=1}^{n} (r_i - 1).q_i$

  - MDL metric $Q_{MDL}(B_S, D)$ defined as

$$Q_{MDL}(B_S, D) = H(B_S, D) + \frac{K}{2} \log N \quad (6)$$

  - Bayesian metric $Q_{Bayes}(B_S, D)$ defined as

$$Q_{Bayes}(B_S, D) = \prod_{i=0}^{n} \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(r_i - 1 + N_{ij})!} \prod_{k=1}^{r_i} N_{ijk}! \quad (7)$$

- **direct acyclic graph (DAG):** the second stage consists in directing all the edges in the skeleton to get a DAG. The first step in directing arrows is to check for every configuration $x - -z - -y$ where $x$ and $y$ not connected in the skeleton whether $z$ is in the set $Z$ of variables that justified removing the link between $x$ and $y$. If $z \notin Z$, we can assign direction $x \rightarrow z \leftarrow y$.

  Then, a set of rules is applied to direct the remaing edges:
  - rule 1: if $i \rightarrow j - -k$ & $i - / - k$ then $j \rightarrow k$
  - rule 2: if $i \rightarrow j \rightarrow k$ & $i - -k$ then $i \rightarrow k$
  - rule 3: if $i \rightarrow j \leftarrow k$ & $S_1$ then $m \rightarrow j$
  - rule 4: if $i \rightarrow j$ & $S_2$ then $i \rightarrow m$ & $k \rightarrow m$
  - rule 5: if no edges are directed then take a random one (first we can find).

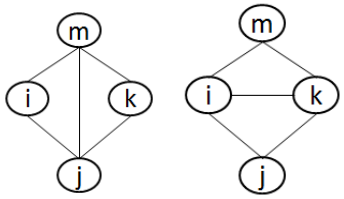


Fig. 3. The two structures $S_1$ (left) and $S_2$ (right)

*2) Decision trees:* Introduced in 1984 by Breiman et *al.* [30], decision trees are one of the few knowledge representation schemes which are easily interpreted and may be inferred by very simple learning algorithms [32]. A decision tree is a tree in which: (i) each internal node tests an attribute, (ii) each branch corresponds to attribute value and (iii) each leaf node assigns a classification. Decision trees are potentially powerful predictors and provide an explicit concept description for a dataset.

Nowadays, several methods have been proposed for constructing decision trees. For this study, we have used the most used methods in the state in the art such as C4.5 [33], CART [30] and BFTREE [34]. We have tested these algorithms on our dataset and we found that C4.5 algorithm outperformed the others. Therefore, we will present a brief description of this decision trees algorithm used in our study.

C4.5 [33] is an extension of the *ID3* [35] algorithm developed by Ross Quinlan in 1986. C4.5 builds decision trees from a set of training data using the concept of information entropy defined in Eq. 8. At each node of the tree, C4.5 chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. Its criterion is based on the gain ratio, defined in Eq. 12, that results from choosing an attribute for splitting the data. The attribute with the highest gain ratio is chosen to make the decision. The C4.5 algorithm then recurses on the smaller sublists.

$$Entropy(p) = -\sum_{k=1}^{c} P(k/p) \times \log(P(k/p)) \tag{8}$$

where:

- $N(p)$ is the cardinality of the set of observations associated to the position $p$ in the database of observations $D$;
- $N(k/p)$ is the cardinality of the set of observations associated to the position $p$ belonging to the class $k$;
- $P(k/p)$ is defined as:

$$P(k/p) = \frac{N(k/p)}{N(p)} \tag{9}$$

$$Gain(p,test) = Entropy(p) - \sum_{i=1}^{n} P_i \times Entropy(p_i) \tag{10}$$

$$SplitInfo(p,test) = -\sum_{i=1}^{n} P\prime(i/p) \times \log(P\prime(i/p)) \tag{11}$$

$$GainRatio(p,test) = \frac{Gain(p,test)}{SplitInfo(p,test)} \tag{12}$$

*3) Performance metrics:* Classifiers are useful tools which are commonly used in decision analysis, to help identifying a strategy most likely to reach a goal [36]. They provide a highly effective and simple structure that can be explored to make predictions and decisions. Despite the obvious advantages of these tools, they do not provide a 100% accuracy result. Due to this inaccuracy, several performance criteria have been proposed in the state of the art [37] to identify the quality of a classifier. The main criteria used are:

- *Accuracy*: denotes the percentage of the correctly classified instances;
- *Area under the ROC curve (AUC)*: it is equal to the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. In our study, AUC is estimated using Mann-Whitney statistic test as presented in [38]. The AUC of a classifier $G$ is defined as:

$$\widehat{AUC} = \frac{S_0 - n_0(n_0+1)/2}{n_0 1} \tag{13}$$

  where $n_0$ and $n_1$ are the numbers of positive and negative examples respectively, and $S_0 = \sum r_i$, where $r_i$ is the rank of the $i$th positive example in the ranked list. The authors in [38] suggest that its use should replace accuracy when measuring and comparing classifiers: the best classifier is the one with the largest $AUC$;
- *comprehensibility*: qualifies the exploitability of the produced model. For example, in a Bayesian network, the important number of a node's parents affects the identification of its strong relations with them;
- *classification rapidity*: which also would be a crucial factor if the training dataset is huge for example.

Since the size of our dataset is not important (less than 100 records, 1 record for each respondent), we have used only the following metrics to determine the best produced model, from the two classifiers, that fits our dataset: *Accuracy, AUC* and *comprehensibility*.

## IV. Experimental results

In this section, we analyze respondents' answers to extract some knowledge about their perception on the two studied systems: keystroke dynamics and face verification systems. In order to perform Kruskall-Wallis tests, Decision Trees and Bayesian networks, we have developed an application that uses the *WEKA* library [31]. In the next section, we present the volunteer crew that participated in this study.

### A. Test protocol

The pilot study was distributed on a paper sheet to a sample of 70 volunteers, including students (71.4%) coming from different countries and employees (28.6%). Tests have been conducted in public places over a 2 months period. It consists in testing both systems (*enrollment* then multiple *verifications* playing the role of an impostor and a legitimate user). Then, they were requested to answer a questionnaire: part A, part B and two times part C (one for each system). Volunteers completed the survey voluntarily and received none remuneration. During the tests, volunteers were informed about the purpose of the study, and their responses would be confidential and anonymous. The age and gender distribution of the volunteer crew are shown in Fig.4.

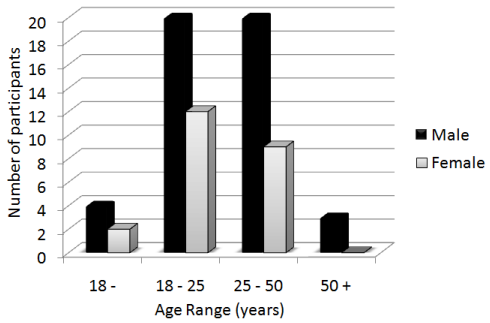


Fig. 4. Age and gender distribution of the volunteer crew

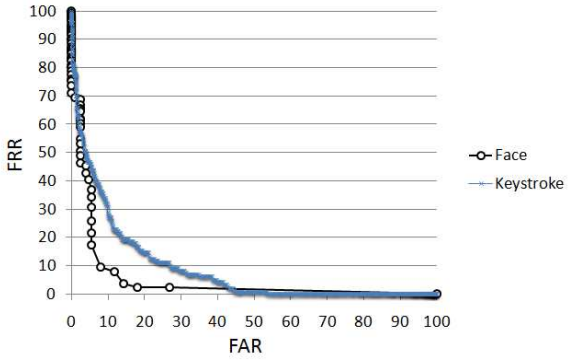


Fig. 5. DET curves for the two tested biometric systems

### B. Test materials

In this study, we have used two biometric systems. Their performances are calculated with captures provided by our volunteer crew. We plot their Detection Error Trade-off (DET) curves in Fig. 5:

- Keystroke verification system: It is a biometric system based on behavioral analysis developed in our research laboratory [15]. The main goals of this software is to allow the creation of a keystroke dynamics database [28] and to compare different algorithms in the state of the art, within the same conditions (e.g., acquisition conditions), for evaluation issues. The system provides an *EER* value equals to 17.51% on a database composed of 70 individuals with 3 vectors used for *enrollment* and 2 for the *tests*. The system implements a score-based method presented in [39]. It is a statistical method based on the average ($\mu$) and standard deviation vectors ($\sigma$) computed with the enrollment vectors, with $v$ a test vector of n dimensions:

$$score = 1 - \frac{1}{n}\sum_{i=1}^{n} exp\left(-\frac{|v_i - \mu_i|}{\sigma_i}\right) \qquad (14)$$

- Face verification system: It is a biometric system based on morphological analysis developed in our research laboratory. The system uses comparisons based on keypoints detection using SIFT descriptors. It provides an *EER* value equals to 8.76% on a database composed of 70 individuals with 1 image used for *enrollment* and 2 for the *tests*.

### C. Data pre-processing

The first step of the proposed methodology consists in the deletion of respondents' answers that did not answer a certain number of questions of the questionnaire. For 3 unanswered questions, two vectors of answers (one from keystroke verification system and the other from face one) have been eliminated from this study. Therefore, the results presented in the next sections are done using 69 vectors of answers on both systems.

### D. Respondent demographics analysis

We study in this section the relationship, on each system, between respondents' demographic characteristics and their answers on perception questions. Table I shows the results for a confidence degree $= 95\%$. Bold values indicate significant relationships based on the criterion defined in Eq. 3. From this table, we can put into obviousness these significantly relationships:

- gender was significantly related to their knowledge about biometric technology: males have expressed more knowledge than females;
- age was significantly related to their answers about keystroke's robustness against attacks: aged respondents ($\geq$ 28) considered that the system is more robust against attacks than youngest ones;
- education level was significantly related to their opinions about secret-based solutions against fraud: high school graduate respondents found that secret-based solutions are less appropriate against fraud than college graduate ones;
- education level was significantly related to the disturb factor while using both systems: none graduated respondents were much more disturbed while using face verification system than the others. For keystroke system, they were much more disturbed than high school graduate respondents;
- education level was significantly related to their concerns about privacy issues while using keystroke system: none graduated respondents have expressed much more concerns about their privacy than the others;
- education level was significantly related to their willingness to use the face system in the future: college graduate respondents have expressed more willingness to use it than the others;
- education level was significantly related to their trust on face system: high school graduate respondents have expressed less trust on it than the college graduate ones.

### E. Comparative study of the studied systems

In this section, we present respondents' knowledge about biometric technology and a comparative analysis between the studied systems based on a statistical analysis of their

TABLE I
PERCEPTION QUESTIONS AND DEMOGRAPHIC FACTORS, KRUSKALL-WALLIS ANALYSIS: LINES WITH TWO P-VALUES CORRESPOND TO SYSTEMS' SPECIFIC QUESTIONS (FACE/KEYSTROKE)

| *Perception questions* | *Gender* | *Age* | *Education* | *Profession* |
|---|---|---|---|---|
| Biometric technology knowledge | **0.01** | 0.86 | 0.21 | 0.73 |
| Awareness about fraud identity | 0.18 | 0.42 | 0.12 | 0.15 |
| Secret-based against fraud | 0.18 | 0.6 | **0.008** | 0.61 |
| Biometric-based against fraud | 0.19 | 0.32 | 0.09 | 0.54 |
| Disturbed | 0.43/0.87 | 0.1/0.91 | **0.02/0.02** | 0.23/0.22 |
| Threats to privacy | 0.51/0.89 | 0.95/0.76 | 0.19/**0.006** | 0.7/0.23 |
| Easy to use | 0.46/0.26 | 0.07/0.31 | 0.47/0.8 | 0.61/0.65 |
| Verification fast | 0.2/0.14 | 0.42/0.85 | 0.26/0.16 | 0.13/0.26 |
| Correct answer | 0.91/0.13 | 0.79/0.44 | 0.13/0.05 | 0.72/0.09 |
| System can be easily attacked | 0.9/0.4 | 0.63/**0.01** | 0.06/0.48 | 0.88/0.08 |
| Use in the future | 0.14/0.14 | 0.46/0.98 | **0.02**/0.74 | 0.38/0.19 |
| Trust | 0.26/0.71 | 0.13/0.59 | **0.04**/0.94 | 0.98/0.11 |
| General appreciation | 0.07/0.12 | 0.2/0.56 | 0.26/0.52 | 0.07/0.58 |

answers. A Kruskall-Wallis test was performed to identify the significant differences among this comparison. Table II shows the results for a confidence degree $= 95\%$. Bold values indicate significant relationships based on the criterion defined in Eq. 3. From the answers given by respondents and table II, we can put into obviousness some interesting points:

- most of the respondents (72.5%) have already heard before our study of biometric authentication systems and less than half of them (43.5%) have already used a biometric system;
- 43.5% of the respondents have expressed a good knowledge about biometric technology;
- using Kruskall-Wallis test (p-value $< 0.01$), respondents considered that biometric technology (92.75% agree) is much more appropriate than secret-based solutions (39.13% agree) against fraud;
- there were no significant differences on disturbed, easy to use, willingness to use the system in the future and trust factors. 21.74% were disturbed while using face system and 13.04% for keystroke one. 11.6% of the respondents have found that face system is not easy to use and 8.7% for keystroke one. 27.54% of the participants hesitate or refuse the use of face system in the future and 15.94% for keystroke one. For their perception about trust, 33.33% do not trust face system and 23.19% for keystroke one;
- there were significant differences about their concerns for privacy issues, their perception about systems performances and their general appreciation among the studied systems. Respondents have expressed much more concerns about their privacy while using face system (46.4%) than keystroke one (13.04%). They found that keystroke performance outperformed face one. For their general appreciation, they were more satisfied from the use of keystroke system (89.85%) than face one (81.16%);
- finally, 26.1% prefer to use the face system and 56.52% for keystroke one for managing logical access, 36.23% prefer to use the face system and 14.5% for keystroke one for physical access, 31.88% prefer to use the face system and 26% for keystroke one for both kinds of access. This indicates that the keystroke system is more requested to be used for managing logical access, while the other one for physical access.

TABLE II
COMPARATIVE ANALYSIS OF PERCEPTION BETWEEN THE STUDIED SYSTEMS, KRUSKALL-WALLIS ANALYSIS

| *Perception questions* | *Face system* | *Keystroke system* | *p-value* |
|---|---|---|---|
| Disturbed | 1.8 | 1.54 | 0.05 |
| Threats to privacy | 2.33 | 1.52 | $<<$ **0.05** |
| Easy to use | 3.41 | 3.39 | 0.96 |
| Verification fast | 3.46 | 3.47 | 0.68 |
| Correct answer | 3.36 | 3.72 | **0.01** |
| System can be easily attacked | 2.6 | 2.38 | 0.22 |
| Use in the future | 3.06 | 3.2 | 0.23 |
| Trust | 2.96 | 3.03 | 0.7 |
| General appreciation | 2.98 | 3.26 | **0.02** |

### F. Discussion

The results of this comparative study and the statistical analysis of answers brought many interesting information. We found a frustrating rate (46.4%) concerning their concerns about privacy issues while using face system. The results also brought surprising rates concerning systems performance and their general appreciation. Respondents found that keystroke's performance outperformed face one and they were more satisfied from the keystroke system than the face one. Therefore, it would be important to explain respondents' answers to more understand these rates and the significant differences among the studied systems. This is what we present in the next section.

### G. Data mining analysis

The purpose of this section is to study the dependences between perception questions, $Q_i \; for \; i = 1:17$, to understand respondents' answers and perception. We would also like to more understand the surprising rates provided by the previous section. Due to the nature of construction of Decision Trees and Bayesian networks, missing values (i.e. questions without answers) are handled for both kinds of attributes. For nominal attributes, they are replaced by the most frequently one. While for numerical ones, they are replaced by the average one. Using Bayesian networks and Decision Trees, several points can be concluded:

- Using table III, respondents' concerns about their privacy while using the face system can be explained by their willingness to use the system in the future and their perception about its robustness against attacks. From the respondents that have expressed concerns about their privacy, 27.6% of them hesitate or refuse its use in the future and 31% found that it can be easily attacked. Since most of the respondents (63.77%) do not found the system robust against attacks and 27.54% hesitate or against its use in the future, this explains why a lot of respondents (46.4%) have expressed such concerns;
- Respondents' perception about keystroke performance was related to their general appreciation and if they felt disturbed while using it (Fig. 7). For the face system, it was related to their trust on the system and if they have already tried before a biometric system (Fig. 6). Since most of the respondents (89.85%) were satisfied

from the keystroke system and most of them (84%) were not disturbed while using it, their perception about the keystroke's performance was important. On the other hand, since 56.52% from the respondents have not already tried a biometric system and 33.33% from them do not trust face system, their perception about the face's performance was not important. These relationships explain why respondents found the keystroke's performance outperformed the face one;

- Respondents' general appreciation on the face system was related to their trust on it (table IV). For the keystroke system, it was related to its performance and their trust on it (Fig. 7). We found that the trust factor was an important one that affects their general appreciation in both systems. Since there was no significant difference of trust factor in both systems, we conclude that their perception about keystroke's performance was the main reason for which they were more satisfied in keystroke system than face one;

TABLE III
EXCERPT FROM DECISION TREE EXPLAINING RESPONDENTS' ANSWERS FOR PRIVACY OF THE FACE VERIFICATION SYSTEM. BOLD RULES INDICATE IMPORTANT ONES.

| |
|---|
| **if (willingness to future use = strongly disagree) then intrusive (2.0)** |
| **if (willingness to future use = disagree) then** |
| ... ... ... |
| **if (fraud awareness $>$ 1) then intrusive (6.0)** |
| ... ... ... |
| if (willingness to future use = agree) then |
| ... ... ... |
| **if (system can be easily attacked $>$ 2) then** |
| if (biometric knowledge $\leq$ 2) then not intrusive (10.0/3.0) |
| if(biometric knowledge $>$ 2) then |
| **if (tried biometric before = yes) then quite intrusive (4.0)** |
| **if (tried biometric before = no) then intrusive (7.0/2.0)** |
| if (willingness to future use = strongly agree) then |
| ... ... ... |
| *Accuracy: 79.71%* |
| AUC: |
| *not at all intrusive: 0.96, not intrusive: 0.91, intrusive:0.95 and quite intrusive: 0.96* |

TABLE IV
EXCERPT FROM DECISION TREE EXPLAINING RESPONDENTS' ANSWERS FOR THEIR GENERAL APPRECIATION OF THE FACE VERIFICATION SYSTEM. BOLD RULES INDICATE IMPORTANT ONES.

| |
|---|
| **if (trust = no at all) then** |
| **if (fraud awareness $\leq$ 2) then not at all satisfied (2.0)** |
| if (fraud awareness $>$ 2) then satisfied (2.0) |
| **if (trust = not really) then** |
| if (correct answer = never) then satisfied (0.0) |
| **if (correct answer = rarely) then not satisfied (1.0)** |
| if (correct answer = sometimes) then |
| **if (biometric appropriate solution $\leq$ 3) then not satisfied (3.0)** |
| ... ... ... |
| if (trust = rather) then satisfied (31.0/6.0) |
| if (trust = yes) then |
| ... ... ... |
| *Accuracy: 85.5%* |
| AUC: |
| *not at all satisfied: 0.9, not satisfied: 0.92, satisfied: 0.82, and quite satisfied: 0.88* |

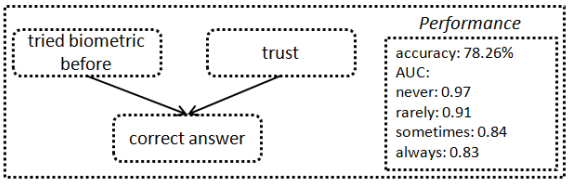


Fig. 6. Excerpt from the bayesian network explaining respondents' answers of performance for face system

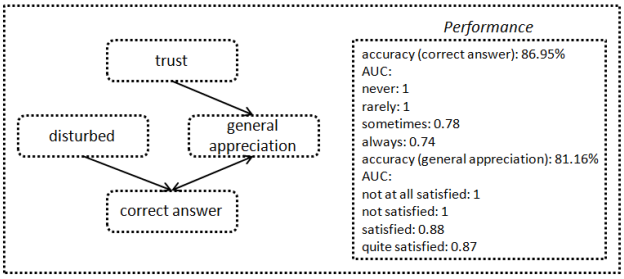


Fig. 7. Excerpt from the bayesian network explaining respondents' answers of performance and general appreciation for keystroke system.

## V. CONCLUSION AND PERSPECTIVES

A complementary evaluation methodology to performance evaluation is proposed in this paper. The proposed methodology focus on studying users' perception while using a biometric system to qualify its operationality and acceptability. It uses (i) Kruskall-Wallis test to determine if there is a significant relationship between demographic characteristics and users' perception and (ii) two data mining approaches, Bayesian networks and Decision Trees, that illustrate the dependencies between respondents' answers to explain their answers and behaviors. We have applied this methodology on 70 persons using two biometric systems based on physical (face verification, $EER = 8.76\%$) and behavioral (keystroke dynamics, $EER = 17.51\%$) analysis. The main results of the survey are:

- respondents considered that biometric-based technology is more appropriate than secret-based solutions against fraud;
- demographic factors (age, gender and education level) have affected their answers on some perception questions;
- both systems were acceptable and respondents were more satisfied with the keystroke system (89.85%) than the other one (81.16%). Trust factor has been identified as a major one that affects their general appreciation on both systems;
- the robustness of the face system against attacks has been identified as an important factor that affects their concerns about privacy issues (46.4%);
- Finally, from the volunteers that they have willingness to use the studied systems in the future, the keystroke system was more requested to be used to manage logical access and the other one for physical access.

Results presented in this paper show that users' perception is a crucial factor that we have to take into account when developing and evaluating biometric systems. Even if the performance of a biometric system outperformed another one, this will not necessarily mean that it will be more operational or acceptable. Robustness of a system against attacks and its perceived trust have been identified as important factors to take into account when designing biometric system. In our point of view, the main drawback of the widespread use of biometric technology is the lack of a *generic evaluation methodology* that evaluates biometric systems taking into account: performance, users' acceptance and satisfaction, data quality and security aspects. We intend to work on data quality and security aspects in the future.

## ACKNOWLEDGMENT

The authors would like to thank the Basse-Normandie Region and the French Research Ministry for their financial support of this work.

## REFERENCES

[1] Y. Chen and A. Jain, "Beyond minutiae: A fingerprint individuality model with pattern, ridge and pore features," in *ICB09*, 2009.

[2] H. Mendez, C. Martin, J. Kittler, Y. Plasencia, and E. Garcia Reyes, "Face recognition with lwir imagery using local binary patterns," in *ICB '09: Proceedings of the Third International Conference on Advances in Biometrics*, 2009.

[3] "Information technology biometric performance testing and reporting," ISO/IEC 19795-1, Tech. Rep., 2006.

[4] M. Theofanos, B. Stanton, and C. A. Wolfson, *Usability & Biometrics: Ensuring Successful Biometric Systems*. National Institute of Standards and Technology (NIST), 2008.

[5] U. Uludag, "Secure biometric systems," Ph.D. dissertation, Michigan State University, 2006.

[6] H. Proença and L. A. Alexandre, "UBIRIS: A noisy iris image database," in *Proceed. of ICIAP 2005 - Intern. Confer. on Image Analysis and Processing*, 2005.

[7] D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain, "Fvc2004: Third fingerprint verification competition," in *in Proceedings of the First International Conference on BiometricAuthentication*, 2004.

[8] D. Petrovska and A. Mayoue, "Description and documentation of the biosecure software library," BioSecure, Tech. Rep., 2007.

[9] E. P. Kukula and R. W. Proctor, "Human-biometric sensor interaction: Impact of training on biometric system and user performance," in *Proceedings of the Symposium on Human Interface 2009 on Human Interface and the Management of Information. Information and Interaction. Part II*, 2009.

[10] A. K. Jain, S. Pankanti, S. Prabhakar, L. Hong, and A. Ross, "Biometrics: A grand challenge," *Pattern Recognition, International Conference*, 2004.

[11] E. P. Kukula, C. R. Blomeke, S. K. Modi, and S. J. Elliott, "Effect of human-biometric sensor interaction on fingerprint matching performance, image quality and minutiae count," *Int. J. Comput. Appl. Technol.*, 2009.

[12] M. Theofanos, B. Stanton, S. Orandi, R. Micheals, and N. Zhang, "Usability testing of ten-print fingerprint capture," National Institute of Standards and Technology (NIST), Tech. Rep., 2007.

[13] M. Theofanos, B. Stanton, C. Sheppard, R. Micheals, N.-F. Zhang, J. Wydler, L. Nadel, and W. Rubin, "Usability testing of height and angles of ten-print fingerprint capture," National Institute of Standards and Technology (NIST), Tech. Rep., 2008.

[14] "Public Attitudes Toward the Uses of Biometric Identification Technologiesby Government and the Private Sector," Opinion Research Corporation International (ORC), Tech. Rep., 2002.

[15] R. Giot, M. El-Abed, and C. Rosenberger, "Keystroke dynamics authentication for collaborative systems," *Collaborative Technologies and Systems, International Symposium on*, 2009.

[16] F. Deane, K. Barrelle, R. Henderson, and D. Mahar, "Perceived acceptability of biometric security systems," *Computers & Security*, 1995.

[17] L. Coventry, A. D. Angeli, and G. Johnson, "Honest it's me! self service verification," in *CHI 2003*, 2003.

[18] S. Schimke, C. Vielhauer, P. Dutta, T. Basu, A. De Rosa, J. Hansen, J. Dittmann, and B. Yegnanarayana, "Cross cultural aspects of biometrics," in *Workshop Proceedings Biometric Challenges arising from Theory to Practice*, 2004.

[19] J. Moody, "Public perceptions of biometric devices: The effect of misinformation on acceptance and use," in *the Informing Science and Information Technology Education*, 2004.

[20] L. A. Jones, A. I. Antón, and J. B. Earp, "Towards understanding user perceptions of authentication technologies," in *Proceedings of the 2007 ACM Workshop on Privacy in the Electronic Society*, 2007.

[21] S. J. Elliott, S. A. Massie, and M. J. Sutton, "The perception of biometric technology: A survey," *Automatic Identification Advanced Technologies*, 2007.

[22] R. R. Heckle, A. S. Patrick, and A. Ozok, "Perception and acceptance of fingerprint biometric technology," in *Proceedings of the 3rd Symposium on Usable Privacy and Security*, 2007.

[23] A. P. Pons and P. Polak, "Understanding user perspectives on biometric technology," *Communications of the Association for Computing Machinery (ACM)*, 2008.

[24] C. Riley, G. Johnson, H. Mccracken, and A. Al-Saffar, "Instruction, feedback and biometrics: The user interface for fingerprint authentication systems," in *INTERACT '09: Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction*, 2009.

[25] L. Coventry, A. De Angeli, and G. Johnson, "Usability and biometric verification at the atm interface," in *CHI '03: Proceedings of the SIGCHI conference on Human factors in computingsystems*, 2003.

[26] "Uk passport service biometrics enrolment trial," Atos Origin, Tech. Rep., 2005.

[27] E. Kukula and S. Elliott, "Implementation of hand geometry at purdue university's recreational center: an analysis of user perspectives and system performance," *39th International IEEE Carnahan Conference on Security Technology*, 2006.

[28] R. Giot, M. El-Abed, and C. Rosenberger, "Greyc keystroke : a benchmark for keystroke dynamics biometric systems," in *IEEE Third International Conference on Biometrics : Theory, Applications and Systems (BTAS)*, 2009.

[29] N. Friedman, D. Geiger, and M. Goldszmidt, "Bayesian network classifiers," *Mach. Learn.*, 1997.

[30] L. Breiman, J. Friedman, R. Olshen, and C. Stone, "Classification and regression trees," Wadsworth International Group, 1984.

[31] R. R. Bouckaert, E. Frank, M. Hall, R. Kirkby, P. Reutemann, A. Seewald, and D. Scuse, "Weka manual," Department of Computing Science, University of Waikato, New Zealand, Tech. Rep., 2009.

[32] J. Baldwin and D. Xie, "Simple fuzzy logic rules based on fuzzy decision tree for classificationand prediction problem," *Intelligent information processing II*, 2005.

[33] J. R. Quinlan, *C4.5: Programs for Machine Learning (Morgan Kaufmann Series in Machine Learning)*. Morgan Kaufmann, 1993.

[34] H. Shi, "Best-first decision tree learning," Master's thesis, University of Waikato, 2006.

[35] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, 1986.

[36] B. Edwards, M. Zatorsky, and R. Nayak, "Clustering and classification of maintenance logs using text data mining," in *Seventh Australasian Data Mining Conference (AusDM 2008)*, 2008.

[37] R. Ricco, "Graphes d'induction," Ph.D. dissertation, Universit Claude Bernard - Lyon 1, 1997.

[38] C. X. Ling, J. Huang, and H. Zhang, "Auc: A better measure than accuracy in comparing learning algorithms," in *Canadian Conference on AI*, 2003.

[39] S. Hocquet, J. Ramel, and H. Cardot, "User classification for keystroke dynamics authentication," in *ICB07*, 2007, pp. 531–539.

## VI. VITA

Mohamad El-Abed is a PhD student in the GREYC laboratory. He obtained his Master of Science in 2008 from the University of Rouen. His research interests biometrics, especially the evaluation of biometric systems.

Romain Giot is a research engineer in the GREYC laboratory. He obtained his Master of Science in 2008 from ENSICAEN. His research interests biometrics, especially the definition of keystroke dynamics biometric systems and multibiometric systems.

Baptiste Hemery is an assistant professor at IUT of Saint-Lo (France). He obtained his Ph.D. from the University of Caen Basse-Normandie in 2009. He belongs to the GREYC laboratory in the computer security research unit. His research interests concern image interpretation evaluation and biometric systems.

Christophe Rosenberger is a Full Professor at ENSICAEN, France. He obtained his Master of Science in 1996 and its Ph.D. degree in 1999 from the University of Rennes I. He works at the GREYC laboratory. His research interests include computer security and biometrics. He is particularly interested in authentication methods for e-transactions applications.

## APPENDIX
## SURVEY QUESTIONNAIRE

| **Part A. Socio-demographic characteristics** | |
|---|---|
| Date of birthday | ... |
| Gender | ☐ male ☐ female |
| In which continent do you live? | ☐ asia ☐ europe ☐ north America ☐ south America ☐ other |
| Highest education level | ☐ high school graduate ☐ college graduate ☐ other |
| Profession | ☐ student ☐ worker ☐ retired ☐ other |

| **Part B. General perception of biometric systems** | |
|---|---|
| $Q_1$. Have you ever heard before about biometric authentication systems **(before our study)**? | ☐ yes ☐ no |
| $Q_2$. Have you ever tried a biometric system **(before our study)**? | ☐ yes ☐ no |
| $Q_3$. Have you ever been personally the victim of identity fraud? | ☐ yes ☐ no |
| $Q_4$. How would you rate your knowledge about biometric technology? | ☐ not at all important ☐ not important ☐ almost important ☐ quite important ☐ I do not know |
| $Q_5$. How would you rate your awareness about fraud identity? | ☐ not at all important ☐ not important ☐ almost important ☐ quite important ☐ I do not know |
| $Q_6$. In your opinion, are secret-based solutions (eg. password) an appropriate solution against fraud (eg. e-commerce)? | ☐ strongly disagree ☐ disagree ☐ agree ☐ strongly agree ☐ I do not know |
| $Q_7$. In your opinion, are biometric-based solutions an appropriate solution against fraud (eg. e-commerce)? | ☐ strongly disagree ☐ disagree ☐ agree ☐ strongly agree ☐ I do not know |

| **Part C. Perception of the tested system** | |
|---|---|
| $Q_8$. Have you ever tried this biometric modality **(before our study)**? | ☐ yes ☐ no |
| $Q_9$. were you disturbed while using this system? | ☐ not at all disturbed ☐ not disturbed ☐ disturbed ☐ quite disturbed ☐ I do not know |
| $Q_{10}$. does this technology threats your privacy? | ☐ not at all intrusive ☐ not intrusive ☐ intrusive ☐ quite intrusive ☐ I do not know |
| $Q_{11}$. is it easy to use this system? | ☐ not at all easy ☐ not easy ☐ easy ☐ quite easy ☐ I do not know |
| $Q_{12}$. Do you find the verification fast? | ☐ not at all fast ☐ not fast ☐ fast ☐ quite fast ☐ I do not know |
| $Q_{13}$. Is the answer of the biometric system is correct? | ☐ never ☐ rarely ☐ sometimes ☐ always ☐ I do not know |
| $Q_{14}$. In your opinion, is the system used can be easily attacked? | ☐ strongly disagree ☐ disagree ☐ agree ☐ strongly agree ☐ I do not know |
| $Q_{15}$. Are you ready to use this biometric system in the future? | ☐ strongly disagree ☐ disagree ☐ agree ☐ strongly agree ☐ I do not know |
| $Q_{16}$. If you are ready to use this system in the future, would you like to use it for physical (eg. access a building) or logical (eg. log on to a computer) access? | ☐ physical ☐ logical |
| $Q_{17}$. do you trust this system? | ☐ no at all ☐ not really ☐ rather ☐ yes ☐ I do not know |
| $Q_{18}$. What is your general appreciation of this system? | ☐ not at all satisfied ☐ not satisfied ☐ satisfied ☐ quite satisfied ☐ I do not know |