UNIVERSIDAD CARLOS III DE MADRID

TESIS DOCTORAL

Evaluation Methodologies for Security Testing of Biometric Systems beyond Technological Evaluation

Autor:
Mª Belén Fernández Saavedra

Director/es:
Raúl Sánchez Reíllo

DEPARTAMENTO DE TECNOLOGÍA ELECTRÓNICA

Leganés, Marzo 2013

PhD Thesis

Evaluation Methodologies for Security Testing of Biometric Systems beyond Technological Evaluation

Author: Mª Belén Fernández Saavedra

Advisor: Raúl Sánchez Reíllo

Signature from the Thesis Committee:

	Name and Surname	Signature
President:	Arturo Ribagorda Garnacho	
Vocal:	Michael Fairhurst	
Secretary:	Carmen Sánchez Ávila	

Mark:

Leganés, de Marzo de 2013

To my favourite sister / A mi hermana favorita

&

To whom it may interest / A quien pueda interesar

Acknowledgements

During the development of this PhD Thesis, I have counted on the help of many people, either from a scientific, emotional or economic perspective. Therefore, I would like to thank to all of them their help and support. Nevertheless and due to this section is dedicated to them, I would like to use the most proper language for each. For this reason, I would like to apologize for mixing languages in advance. Moreover, I hope I do not forget anybody. Just in case, sorry for that and thank you very much.

Para empezar quiero darle las gracias a Raúl, mi director de tesis, no sólo por todo lo que me ha enseñado durante estos años, por su tiempo y su dedicación si no sobre todo por su confianza. Mil gracias Raúl por creer en mí más de lo que yo misma lo hago, por ayudarme a superar mis miedos y por qué no, por empeñarte en que pidiera aquella beca para hacer la tesis.

Y por supuesto a también quiero darle las gracias a mis padres Martín y Milagros y a mi hermana Rocío. Muchas gracias por vuestro cariño, por vuestro apoyo incondicional y por haberme ayudado a enfrentarme y superar todos y cada uno de los obstáculos que ha habido en el camino, no solo ahora sino desde siempre. Especialmente quiero agradeceros el estar a mi lado en esta última etapa que ha sido la más difícil. Gracias a los tres por vuestra compresión, por escucharme, por vuestros ánimos y por ayudarme con las pequeñas cosas de cada día que han hecho que yo pudiera dedicarle mucho más tiempo a terminar la tesis.

Thank you to Alexander Nouak and Steve Elliott for giving me the opportunity to work as part of their research groups respectively. It was a pleasure to stay there. Thank you very much for supporting me, for your time and for dealing with the Spanish paperwork. Specially, I would like to say "Danke Schön" to Alexander and, by extension, Christoph Bush and the rest of people of the laboratory for helping me during my first steps in planning a biometric performance evaluation. Likewise, I also would like to thank to Steve and the people of BSPA Lab for clarifying my ideas about usability and HBSI.

Thank you to all experts of ISO/IEC JTC1 SC37 WG5 and to all delegations for commenting and proposing contributions to the ISO/IEC 29197 project. In particular, I would like to thank Rick Lazarick for sharing his thoughts and supporting me with the definition of most proper "environmental terms". Thanks a million.

Además quiero darles las gracias a Marino Tapiador, Miguel Bañón y Jose Emilio Rico por hacer que los Common Criteria parezcan fáciles. Gracias por lo que me habéis enseñado, por resolver mis dudas y por prestarme vuestro apoyo en la difícil tarea de interpretar Common Criteria en el caso de la biometría.

También quiero agradecer a Oscar, Judith y Alberto los consejos y los ánimos que me han dado. Todos habéis pasado por lo mismo antes que yo y vuestra ayuda ha sido vital especialmente cuando tuve que enfrentarme a la dura tarea de escribir. Muchas gracias chicos

por soportar mis nervios y por hacer de "asesores" ya no sólo para la tesis, sino también para los congresos y los artículos.

Gracias a todos los miembros del GUTI y a los que estáis y a los que ya habéis dejado la universidad, porque vuestra ayuda ha sido fundamental a la hora de realizar la tesis y en especial las evaluaciones. Gracias por vuestro trabajo, por hacer de operadores, de usuarios, por montar y desmontar escenarios, por programar lo que hiciera falta, por reclutar usuarios, pero sobre todo por vuestro tiempo, por vuestra paciencia y en especial, por aguantarme. Muchas gracias a RAM, Pelos, Aitor, Inma, Jauma, Ramón, Sandra, Pepe, Jorge, Eugenio, Luis, Michael y Mª Jesús. Y por supuesto a Reillo, Oscar y Judith que también tenéis que estar aquí. Asimismo, también os quiero dar las gracias no sólo a vosotros sino también a los que siempre nos han acompañado, es decir, a los ya conocimos como ya los acólitos del GUTI. Han sido muchísimos los momentos que hemos compartido, los cumpleaños, los viajes, las GUTIcenas, las GUTIbbqs, el GUTIvolley, etc y todos ellos han hecho que los momentos difíciles fueran más llevaderos. Espero que sigan. Por eso también quiero dar las gracias a Paloma, Miguel y Laura, Noe y Daniel, Lourdes, Nando, Mele, Chema, Isa, Juancar, Ivanga, Nerea, Alberto y a alguno más que seguro que me estoy dejando.

Gracias a toda mi familia por su cariño y por darme fuerza para seguir adelante. A mi tío Juan Antonio por sus consejos, sus enseñanzas sobre la vida y sus chascarrillos. Tito tienes razón, esto de la tesis no da dinero. A mi tío Joaquín por apoyo y por sus ánimos con los estudios. Joaquín siento que la tesis no esté en español. Y también a mis primas Carmen y Chave por su afecto y sus palabras de ánimo. A su vez, también quiero darle las gracias a mis abuelas Milagros y Martina y a mi tía Cali. Aunque ya no estéis habéis sido y seguís siendo parte importante de la familia. Creo que ninguno de nosotros os podremos olvidar. Desde aquí os mando un beso muy fuerte.

Gracias a las "Viudas Negras" mi equipo de voleibol, al "Mister" y a los respectivos "esposos". Es inevitable reconocer que sois más que un equipo de voleibol. Gracias por vuestras risas, por los chistes y sobre por esos momentos "viudas" que son inolvidables. Muchas gracias a las Patris, las López incluida Inesita, las Belenes, Leti, Nata, Sara, Miri, las Evas, Eli, Hoze, Javi, Juan y Miguel. Chichas hay que ganar la liga. Miri tienes que rematar tu tesis igual que lo haces en la pista.

Gracias a "mi gente", a Evita y Lucho, Carol y Héctor, Vanesa, Jose y sus niños, Julio, Esther y Javi, Susi, Rocío Núñez y Félix, Miri, Andrés y sus niños, Kris, Nuria, Jorge y Edu. Gracias chicos, grandes y pequeños, por todos los buenos ratos que hemos pasado juntos y que espero sigan pasando. Gracias por apoyarme y animarme y sobre todo porque sé que realmente os da igual que sea una tesis o cualquier otra cosa, habríais hecho lo mismo. Chicos esto ya se acaba, ya dejaré de hablar de tesis.

Gracias a Dori, Tea, Peque y a todos mis peluches. Sí, aunque seáis peluches sois parte de mi vida. Hay gente que tiene perro o gato y yo os tengo a vosotros. Gracias por haberme acompañado durante todo el tiempo de la tesis y especialmente porque por algún extraño

motivo me dais paz y tranquilidad. Bueno, y sobre todo gracias a todos los que me los habéis regalado. Ni que decir tiene que por favor, no me regaléis más.

También quiero agradecer su participación en las evaluaciones a todos los "test subjects" o "usuarios". Vosotros habéis sido una parte importante de esta tesis. Está claro que habéis dejado "huella". Gracias por vuestro tiempo y por vuestra paciencia.

Y como no, gracias al Ministerio de Educación por concederme la beca FPU y subvencionar mis estudios de doctorado. Sin ella, quien sabe si lo hubiera pasado. Aunque fundamentalmente quiero dar las gracias a todos los españoles que con sus impuestos son los que hacen posible que existan este tipo de ayudas.

Por último solo me queda decir ¡¡¡GRACIAS A DIOS!!!

Abstract

The main objective of this PhD Thesis is the specification of formal evaluation methodologies for testing the security level achieved by biometric systems when these are working under specific contour conditions. This analysis is conducted through the calculation of the basic technical biometric system performance and its possible variations. To that end, the next two relevant contributions have been developed.

The first contribution is the definition of two independent biometric performance evaluation methodologies for analysing and quantifying the influence of environmental conditions and human factors respectively. From the very beginning it has been claimed and demonstrated that these two contour conditions are the most significant parameters that may affect negatively the biometric performance. Nevertheless, in spite of ISO/IEC 19795 standard [ISO'06b], which addresses biometric performance testing and reporting, being published in 2006, no evaluation methodology for assessing such adverse effects has been implemented yet. Therefore, this dissertation proposes both methodologies which have been defined in accordance to the following requirements:

- should be general and modality independent for covering the analysis of all kind of biometric systems;
- should conform to the principles and requirements already defined in ISO/IEC 19795 multipart standard; and
- should provide requirements and procedures to accurately define the evaluation conditions to be tested, conduct reproducible test methods and obtain objective and intercomparable results.

The second relevant contribution is the development of detailed guidelines for addressing how to conduct biometric performance evaluations in compliance with Common Criteria [CC]. Common Criteria is currently the only international recognised evaluation framework with which developers have to analyse and demonstrate the level of security achieved by their products. However, the applicability of this methodology to biometrics needs the specification of supplementary guidelines. As a consequence, this dissertation proposes such guidelines which have been specified according to the following requirements:

- should be independent of any biometric modality;
- should be based on previous works published in this topic BTSE [BTSE'01], BEM [BEM'02] and the ISO/IEC 19792 international standard which addresses security evaluation of biometric system;
- should conform to the last version of both Common Criteria and the ISO/IEC 19795 multipart standards; and
- should cover those kinds of biometric performance evaluations that can be repeatable, i.e. technology and scenario evaluations as well as the Common Criteria evaluation activities involved in the execution of such test procedures.

As for the evaluation of the security of biometric systems there is the need of determine their performance, and as such performance also depends on contour conditions, both evaluation methodologies (i.e. environmental and human factors) and Common Criteria guidelines, are merged in order to provide improved evaluation methodology for the security of biometric systems.

Resumen

El objetivo principal de esta Tesis Doctoral es la especificación de metodologías de evaluación formales para analizar el nivel de seguridad alcanzado por los sistemas biométricos cuando estos se encuentran trabajando bajo condiciones de contorno específicas. Este análisis se realiza a través del cálculo del rendimiento técnico básico del sistema biométrico y sus posibles variaciones. A tal efecto, se han elaborado las siguientes contribuciones.

En primer lugar, se han especificado dos metodologías de evaluación de rendimiento biométrico de manera independiente para analizar y cuantificar la influencia de las condiciones ambientales y los factores humanos, respectivamente. Desde los primeros estudios sobre rendimiento biométrico, se ha afirmado y demostrado que éstos son los parámetros más significativos que pueden afectar negativamente al rendimiento biométrico. No obstante, a pesar de que la norma ISO/IEC 19795 que regula la evaluación y documentación del rendimiento de los sistemas biométricos fue publicada en 2006, ninguna metodología que evalúe dichos efectos adversos ha sido implementada hasta el momento. Por lo tanto la presente Tesis Doctoral propone ambas metodologías, las cuáles han sido definidas conforme a las siguientes condiciones:

- son de carácter general e independientes de cualquier modalidad biométrica para cubrir el análisis de todo tipo de sistemas biométricos,
- cumplen con los principios y requisitos previamente definidos en la norma internacional ISO/IEC 19795 [ISO'06b], y
- proporcionan requisitos y procedimientos detallados para: definir las condiciones de los ensayos, efectuar métodos de ensayo reproducibles y obtener resultados objetivos e intercomparables.

En segundo lugar, se han desarrollado directrices específicas que abordan la forma de realizar evaluaciones de rendimiento biométrico conforme a "Common Criteria for IT security evaluation" (conocido habitualmente como "Common Criteria" [CC]). Common Criteria es actualmente el único marco de evaluación internacionalmente reconocido del que disponen los desarrolladores de sistemas biométricos para analizar y demostrar el nivel de seguridad que alcanzan sus productos. Sin embargo, la aplicación de esta metodología a la tecnología biométrica requiere la especificación de pautas complementarias. Por consiguiente, esta Tesis Doctoral propone tales pautas o directrices, las cuáles se han especificado de acuerdo con los siguientes requisitos:

- son independientes de cualquier modalidad biométrica,
- se basan en los trabajos previos que ya han sido publicados en esta área tales como BTSE [BTSE'01], BEM [BEM'02] y el estándar internacional ISO/IEC 19792 [ISO'09a] que regula la evaluación de seguridad de los sistemas biométricos,
- son conformes a las últimas versiones tanto de Common Criteria como de la norma internacional ISO/IEC 19795, y

- cubren tanto el tipo de evaluaciones de rendimiento biométrico que pueden ser repetibles, es decir las evaluaciones tecnológicas y de escenario, como las actividades de evaluación establecidas por la norma Common Criteria que conllevan la realización de dichos procedimientos de test.

Debido a que es necesario determinar el rendimiento de los sistemas biométricos para evaluar su seguridad, y ya que dicho rendimiento depende de distintas condiciones de contorno, las dos metodologías de evaluación previamente definidas (condiciones ambientales y factores humanos) se han unido con las directrices de Common Criteria, para así conseguir una mejora sustancial en la metodología de evaluación de la seguridad de los sistemas biométricos.

Contents

List of Figures

List of Tables

Acronyms

ACO	Composition class
ADV	Development class
AGD	Guidance documents class
ALC	Life-Cycle support class
APE	Protection Profile Evaluation class
ATE	Tests class
ASE	Security target evaluation class
AVA	Vulnerability assessment class
BTSE	Biometric Technology Security Evaluation under the Common Criteria
BEM	Biometric Evaluation Methodology
CC	Common Criteria for Information Technology Security Evaluation
CEN	European Committee for Standardization
CEM	Common Methodology for Information Technology Security Evaluation
CI	Concealed Interactions
CMC	Cumulative Match Characteristic
DET	Detection Error Trade-off
DI	Defective Interactions
EAL	Evaluation Assurance Level
FAR	False Accept Rate
FI	False Interactions
FMR	False Match Rate
FNIR	False Negative Identification Rate
FNMR	False Non-Match Rate
FPIR	False Positive Identification Rate
FRR	False Reject Rate
FTA	Failure To Acquire rate
FTD	Failure To Detect
FTE	Failure To Enrol rate
FTP	Failure To Process
FTX	Failure To Extract
GFAR	Generalized False Accept Rate
GFRR	Generalized False Reject Rate
GUI	Graphical User Interface
H-B	Human-Biometric system interaction
HBSI	Human-Biometric Sensor Interaction
ID	Identifier
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
IT	Information Technology
ITSEF	IT Security Evaluation Facility
NIST	National Institute of Standards and Technology

REC	Reference Evaluation Conditions
REE	Reference Evaluation Environment
ROC	Receiver Operating Characteristic
SAR	Security Assurance Requirement
SAS	Successful Acquisition Sample
SPS	Successful Processed Sample
SDK	Software Development Kit
SFR	Security Functional Requirement
SPD	Security Problem Definition
ST	Security Target
TEC	Target Evaluation Conditions
TEE	Target Evaluation Environment
TOE	Target of Evaluation
TSF	TOE Security Functionality
TSFI	TSF Interfaces

Chapter 1

Introduction

In a short period of time, biometrics has become one of the most relevant technologies used in Information Technology (IT) security. This technology not only provides a mechanism for the protection of assets, but also guarantees that the individual who wants to gain access to them is the real authorized person. This is due to the fact that biometrics consists of the automatic recognition of individuals by analyzing intrinsic human being characteristics which cannot be easily forgotten, lost, exchange or stolen, as it may happen with passwords or cards. Thanks to this property, during the last decade, biometric recognition has been considered the most suitable solution for applications which entails security authentication such as access control, border control, banking, etc. Nevertheless, although biometrics technology is more reliable for people recognition than other IT identification technologies such as tokens or passwords, this technology has two inherent vulnerabilities.

Biometrics is a non-deterministic technology. The recognition process is based on the comparison of biometric samples which is subject to errors. These errors occur due to factors such as the distinctiveness of the biometric characteristic and the repeatability of the acquired biometric samples. Such errors establish the probability that users are correctly recognized or not, determining the accuracy of biometric recognition mechanisms. Both probability measurements are the most important error rates, which, together with acquisition and time metrics, are used to define the technical performance of a biometric system or application. The security strength of biometric technology is quantified through biometric performance and the corresponding error rates, among other parameters.

Moreover, biometric performance is strongly dependent of contour conditions such as users, their interactions and/or the scenario environment. These contour conditions have a significant impact in the sample acquisition process resulting in different negative effects. Such effects cover a huge variety of cases, from biometric sample cannot be captured or it is captured with a deficient quality, to the impossibility of obtaining reliable feature vectors. The effects occurring, as well as their significance, will depend on the specific application as well as on the properties of the particular biometric system such as: modality, the capture device and its acquisition technology, or the implementation of segmentation, feature extraction or quality algorithms. Whatever the case is, the fact is that biometrics is sensitive to all of these conditions and it may cause a significant reduction of biometric performance.

Therefore, in order to analyse the fundamental security of a biometric product, it is necessary to conduct biometric performance evaluations. However, this kind of tests is not straightforward. Due to the probabilistic nature of biometrics, it is indispensable to execute the recognition process several times to achieve statistically significant results. To accomplish this, a considerable number of biometric samples are needed and experiments including genuine and impostor trials must be carried out. In addition, due to the influence of contour conditions, specific testing requirements have to be followed for ensuring similar effects of influential variables in order to avoid biased results. Considering all of these circumstances, a biometric performance evaluation requires the specification of an evaluation methodology which not only defines reproducible test methods, but also obtains intercomparable and reliable results.

In the field of biometrics, there was not any concrete performance evaluation methodology till 2000. Until then, each institution carried out their own evaluations and most of the above mentioned aspects were not considered. As a result, error rates were not trustworthy and most biometric products did not work as well as vendors claimed. It was in that year when the first formal methodology was published and it turned out to become, in 2006, the international standard ISO/IEC 19795-1, Biometric Performance Testing and Reporting –Part 1: Principles and framework [ISO'06b]. Since then, this standard has been expanded with additional parts addressing specific evaluation types or modality considerations. Moreover, several biometric performance evaluations have been conducted with such methodology by a variety of institutions such as private companies, government organizations, universities and independent laboratories. At present, it is a consolidated evaluation methodology and there are laboratories which have been accredited for performing biometric testing in accordance with it. But although ISO/IEC 19795 multipart standard has been defined to obtain biometric performance, evaluation methodologies for testing the influence of contour conditions have not been established yet. Consequently, biometric products cannot be tested properly and elementary factors, such as ambient conditions or user interaction may be unidentified factors that can lead to consider a biometric solution useless from the security point of view.

Likewise, in the field of security an explicit methodology for testing the security of biometric technology did not exist. During the eighties and the early nineties, USA, Canada and

Europe have their own security evaluation methodologies for IT products. However, these methodologies were defined in general terms for covering a wide range of IT products. In 1996, these methodologies were merged into the first version of the so-called Common Criteria (CC) for IT Security Evaluation [CC-1'96]. Throughout the years, this three-part standard together with its Common Methodology (CEM) for IT Security Evaluation have been improved and new versions have been published [CC]. Nevertheless, in spite of the development of new versions, the methodology is still general and does not detail key requirements to carry out biometric performance evaluations appropriately. In different occasions, experts had tried to solve this gap providing supplement guidelines to CC and CEM, or even developing a new international standard, i.e. ISO/IEC 19792:2009, Security techniques – Security evaluation of biometrics [ISO'09a]. But neither the supplement guidelines were accepted by Common Criteria community nor biometric community considers rigorous and well defined the ISO/IEC 19792 standard. Therefore, security evaluations continue being unspecified for testing biometric systems and applications, even though the use of biometrics products is increased more and more.

This dissertation is focused on the development of evaluation methodologies to quantify the effects of contour conditions on biometric system performance as well as the formalization of these methodologies according to the current security evaluation methodologies CC and CEM. The intention is not only to provide the proper procedures to determine the existence of critical factors that affect the basic security of biometric systems and applications, but also to allow that biometric systems will be accurately tested following the CC certification scheme in a similar way than the rest of IT products.

To fulfil the first objective, two methodologies will be specified for analyzing the influence of two of the most relevant factors which affect biometric systems, i.e. environmental conditions and user interaction conditions. These methodologies will be general and modality independent for covering the analysis of all kind of biometric systems. Besides, these methodologies will be defined to conform to the principles and requirements already defined on the ISO/IEC 19795 standard, as well as to be reproducible.

For accomplishing the second objective, specific guidelines will be defined to establish how to conduct biometric performance evaluations and how to interpret contour conditions and their influence on biometric systems as part of a CC evaluation. Particularly, those works units of CEM involved in biometric performance evaluations will be explained in compliance to ISO/IEC 19795 test procedures. Likewise, certain considerations about environment and user's behaviour influence on biometric systems will be detailed from a CC point of view. Moreover, both will be based on the advantageous aspects of previous works and will try to settle those controversial points.

In order to describe the proposed work, this document has been divided in different chapters. Each chapter deals with a specific subject according to the document structure described as follows:

Initially, the three first chapters introduce biometric technology and its evaluation. Specifically, the latter will be described in depth considering both biometric and security perspectives due to the importance for this dissertation. Exactly, these chapters are the following:

- Chapter 2 "Introduction to biometrics": This chapter provides the definition of "biometrics" term and an overview of the biometric technology including existing modalities, the explanation of the general model of a biometric system, as well as the description of the biometric functions.
- Chapter 3 "Evaluation of biometric technology": This chapter explains the concept of biometric testing and offers a taxonomy of the types of biometric evaluations. It also describes in detail the biometric performance evaluation. This description covers what does this kind of evaluation consists of and provides a review of the literature about it, its standardization and the evaluations already carried out.
- Chapter 4 "Security evaluations of biometrics": In this chapter, biometric evaluations are presented from the security point of view. For this purpose, firstly, Common Criteria and its evaluation model are defined. Then, the application of this type of evaluation to biometric system is explained including the existing works.

Then, the following two chapters cover the first objective of this dissertation, as mentioned above. Specifically, these chapters are:

- Chapter 5 "Evaluation methodology for environmental testing of biometric systems": In this chapter, the evaluation methodology for analysing the influence of environmental conditions in biometric performance is established. This evaluation methodology includes the specification of environmental conditions to analyse as well as those requirements for generating, controlling and recording such conditions. Furthermore, the necessary procedures to be carried out during a biometric performance evaluation are described. Such procedures will be requested in addition to the corresponding metrics and measurements to quantify biometric performance and its variations. Finally, experiments executed to validate this methodology as well as the obtained results are shown.
- Chapter 6 "Evaluation methodology for human-biometric system interaction": This chapter establishes the evaluation methodology for analysing the influence of the interaction between the user and the biometric system. This specification entails the definition of all potential conditions to analyse as well as the necessary requirements for studying each aspect possible. Moreover, in the same way that the previous methodology, metrics and measurements to quantify biometric performance and its variations are described, as well as the experiments executed and the results obtained.

After that, the next chapter is focused on the second objective of this dissertation which is the formalization of biometric performance evaluation methodologies according to CC and CEM. In particular, this chapter discusses the following:

- Chapter 7 "Guidelines for conducting biometric performance testing according to CC and CEM": This chapter provides additional guidelines to CEM for applying biometric performance testing methodologies in the context of CC. Besides, it also addresses relevant considerations about contour conditions that affect biometric systems and how to interpret them in terms of CC.

Finally, the last chapter presents the most relevant conclusions obtained throughout this dissertation as well as the research work that can be carried out in the future. Specifically, this chapter is:

- Chapter 8 "Conclusions and future work lines": This chapter summarizes the main conclusions of the work conducted as part of this dissertation and mentions those open research lines that have been identified during its development but whose discussion is out of the scope of this PhD Thesis.

Chapter 2

Introduction to biometric technology

Nowadays, within the context of this work, *biometrics* refers to the science of establishing the identity of a person based on the physical or behavioural attributes associated with an individual [JAI'07, LI'09]. From the practical application of the scientific knowledge in this area, the biometric recognition technology emerges. This chapter presents an overview of this technology.

On one hand, the essential concepts to understand the basis of biometrics science and its technology are given. It includes a detailed explanation of biometrics as well as a list of the most significant properties of the attributes used for the recognition (also known as biometric characteristics or traits). Moreover, depending on such biometric characteristics, different modalities can be distinguished. This chapter also describes the most important biometric modalities and their classification.

On the other hand, this chapter introduces biometric technology from a technical point of view considering both biometric systems and its functionality. For this purpose, a general model of biometric systems including its different components and their interactions are explained. Furthermore, biometric functions to complete the recognition process are provided covering the purpose of enrolment and recognition (verification and identification) phases in addition to the tasks that involve each of them.

2.1 Biometrics

Biometrics is a term derived from the Greek words "bio" (life) and "metron" (measure) and it refers to the statistical analysis of biological observations and phenomena [NSTC'06c, IEEE'09c]. Nevertheless, in the last decades, this term has been also used as an abbreviation of "biometric recognition" in certain fields such as physical and information security and authentication [WAY'00, JAI'07, ISO'07b]. Considering this area, currently biometrics has a more specific definition such as the automated recognition of individuals based on their behavioural and biological characteristics [WAY'00, ISO'07a, DUN'09, LI'09].

The fundamentals of this technology lay in the automatic nature of this process together with the properties of these behavioural and biological characteristics. There are other technologies that allow the automatic recognition of individuals such as ID tokens or passwords, but these technologies entail either that users must have with them a token or that users must memorize a password respectively. Biometrics only requires an intrinsic characteristic of the user. However, to consider it as a biometric characteristic, this should have the following properties [JAIN'98, IEEE'09c]:

- Universality: every individual should have it.
- Uniqueness/distinctiveness: this characteristic should be different across individuals.
- Permanence/robustness: the biometric characteristic should be invariant with time.
- Collectability: it should be possible to acquire and process the characteristic for extracting relevant features without causing any damage to individuals.
- Performance: the level of accuracy at the recognition process using this characteristic should be satisfactory.
- Acceptability: people should be willing to use the system when they have to present such characteristic to the biometric system.
- Circumvention: it should be difficult to imitate or mimic the biometric characteristic in order to avoid its fraudulent usage.

Since biometrics arises as a new recognition technology, several characteristics that meet the above properties to a greater or lesser extent have been discovered. In turn, each of these characteristics has provoked the emergence of different biometric modalities. An overview of these modalities is described in the next section.

2.2 Biometric modalities

Depending on the biometric characteristic used in the recognition process, a wide range of techniques for recognizing individuals exists. Formally, each of these techniques is named biometric modality. Considering a preliminary classification, these modalities can be divided in two main groups [NSTC'06b, IEEE'09a]:

- Physical modalities (also named static or passive). These modalities are based on anatomical or physiological characteristics. Such characteristics are obtained without

the necessity that users perform any specific action. The most common modalities that belong to this group are: fingerprint, face, iris, retina, hand/finger geometry, palm print, vascular pattern recognition and DNA. There are also new modalities such as ear shape or body odour.

- Behavioural modalities (also named dynamic or active): These modalities are based on biometric characteristics that involve the execution of certain activity. Such activity entails a behaviour which has been learned or acquired over time. These modalities are dynamic signature, keystroke, and one of the most recent, gait recognition. Speaker recognition is also another biometric modality that might be classified in this group, although it really involves physical and behavioural features.

The existence of a wide number of biometric modalities as well as their possible combinations cause that there are multiple types of biometric systems. Each of them is implemented with the appropriate biometric capture device/s and algorithm/s to acquire and process the corresponding biometric characteristic/s. However, all of them perform similar operations and have the same components. The following section will explain in detail these systems.

2.3 Biometric systems

For the purpose of biometric recognition, numerous biometric systems have been developed. These systems are responsible for obtaining the necessary user information to accomplish the identification. As it was above mentioned, in spite of the fact that there are several types of biometric systems, they have elements and functions in common. This section presents the general biometric system and its functions.

2.3.1 General biometric system

Typically, every biometric system has the following subsystems: data capture, transmission, signal/image processing, data storage, comparison, decision and administration [IEEE'09c, DUN'09, ISO'10a]. Depending on the specific implementation of the biometric system, some elements may not exist or may not correspond with hardware and software parts. Nevertheless, a common biometric system structure has been established by the biometric community consensus at the International Standard ISO/IEC Standing Document 11 [ISO'10a]. The diagram of this structure is shown in Figure 1. For the purpose of this dissertation, this is the biometric system that is going to consider from now onward.

As it can be seen at the diagram, each subsystem contributes to the recognition process carrying out a particular task. These subsystems and its functionality are described below considering [ISO'07b, IEEE'09d, ISO'10a].

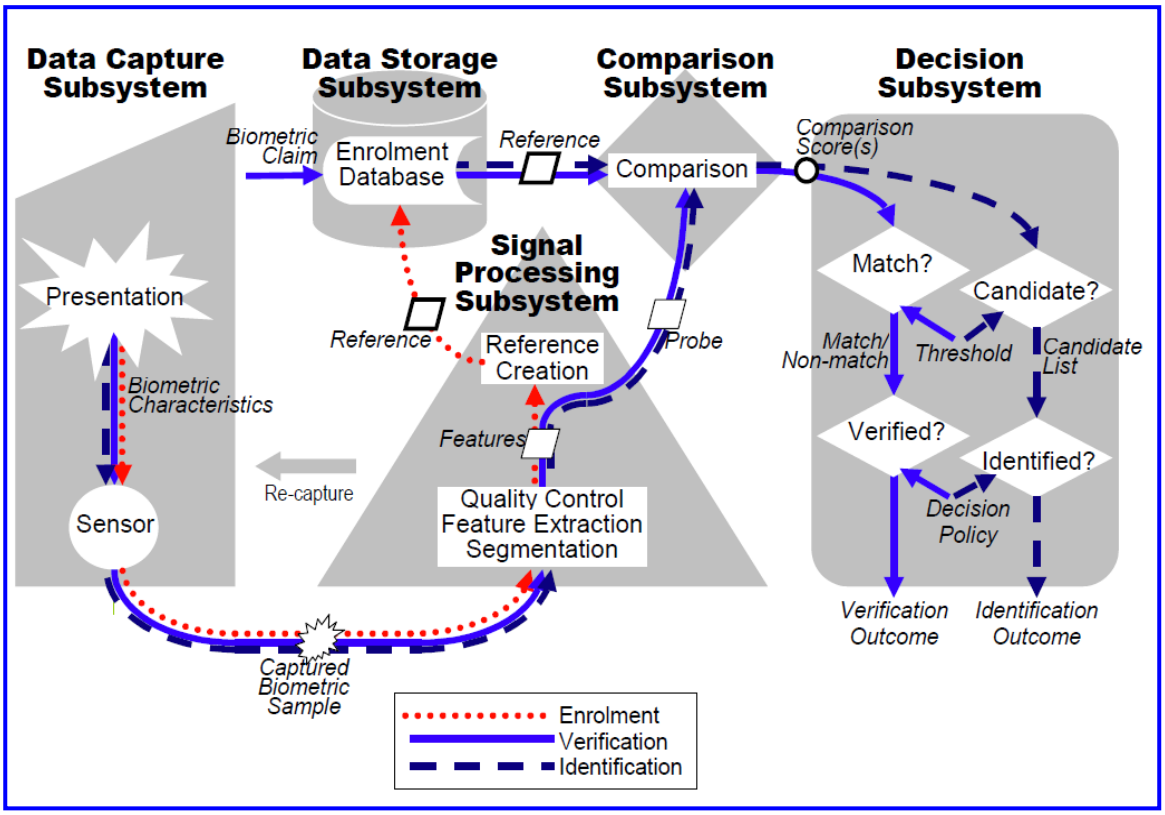

Figure 1. General block diagram for a biometric system [ISO'10a]

2.3.1.1 Data capture subsystem

This subsystem takes part at the beginning of any biometric function. It captures the image or signal which contains the user biometric characteristic and turns it into a digital format for further processing. Such representation of the biometric characteristic is called biometric sample.

Basically, the data capture subsystem is composed of the biometric capture device. This device will be different depending on the biometric modality utilized in the recognition process. For example: a camera is used in iris and face recognition, speaker recognition uses a microphone, keystroke uses a keyboard, etc. In addition, the same type of device can be based on different technologies (i.e. a fingerprint scanner can be a touch or a swipe sensor and both can use different sensing technologies such as optical or capacitive [MAL'09].

2.3.1.2 Signal processing subsystem

Once a biometric sample has been captured, this is sent to the signal processing subsystem. This subsystem is in charge of generating a features vector from the biometric sample. To that end, this subsystem carries out several phases. These are the following:

- Segmentation. During this phase, the useful information of the captured sample is localized and got ready for processing it in the next phases whereas the rest of

information is discarded. For this purpose, this phases includes activities such as detection, alignment, segmentation itself, normalization and enhancement [IEEE'09d].

- Feature extraction. In this phase, the essential features that allow recognizing persons are extracted. For doing that, the information provided by the previous phase is processed using the biometric algorithm. As a result, a representation of the features is generated.
- Quality control. This phase checks if the biometric captured sample, its segmentation or the obtained features satisfy a predefined set of quality specifications. The goal is to detect in advance whether there is any indication that the processing of that sample can cause a failure during the recognition process or reduce the biometric system performance. Depending on the particular analysis to carry out, quality control methods may be applied before or after segmentation and/or before or after feature extraction.

When successful biometric features have been obtained, these will be sent to different subsystems according to the biometric function that it is being executed. In case of the enrolment function, the signal processing subsystem creates a biometric reference from the features. Such reference is sent out to the data storage subsystem. In case of verification or identification functions, the features are sent out to the comparison subsystem directly as a biometric probe. All this is explained in the next sections.

2.3.1.3 Data storage subsystem

This subsystem serves to store biometric references that come from enrolled users. Depending on the specific system it can be a centralized database, a distributed database (either in a personal computer in a local server or in a storage unit of the biometric system itself), or portable device such as a smart card or any ID token. At the same time, personal data from users can be stored together with biometric data.

2.3.1.4 Comparison subsystem

The comparison subsystem compares a feature vector to a single biometric reference in the case of a verification process, or several biometric references stored in the data storage subsystem, in the case of an identification process. The number of biometric references to compare depends on the type of biometric recognition process: verification or identification. Both processes will be explained in section 2.3.2.2. As a result of the comparison process, one similarity score is obtained in case of verification and similarity scores for a list of candidates in case of identification.

2.3.1.5 Decision subsystem

After the comparison process, the result is sent to the decision subsystem. Based on this result and on the decision thresholds specified for the biometric system, this subsystem decides the final result of the recognition process. For verification systems, this result will be to accept or reject the user who claims his/her identity. For identification systems, this result will

be a candidate list which contains the users' identifiers for those users whose biometric references match the biometric sample. This list may be an empty list or a list with a fix number of users' identifiers.

2.3.1.6 Transmission subsystem

This subsystem is composed by all connections between the different parts of the biometric system. It transmits the necessary inputs and outputs between subsystems in order to accomplish all types of biometric functions. Figure 1 illustrates these interactions using arrows of different styles for each of these biometric functions.

2.3.1.7 Administration subsystem

Administration subsystem is a subsystem which is not portrayed in Figure 1 but it is found in most biometric systems. This manages all policies related to the usage of the biometric system. It entails a lot of activities but the most important is controlling the security settings of the biometric system such as quality thresholds, decision thresholds, maximum number of attempts, maximum number of identifiers for the candidate list, etc.

2.3.2 Biometric functions

A biometric system is designed with the purpose of recognizing individuals. This process is composed by two main functions: enrolment and recognition itself. The description of both functions is provided in the following sections.

2.3.2.1 Enrolment

Enrolment function entails the first step of the recognition process. It consists of generating the biometric reference for a person from his/her biometric characteristic and saving it for further comparisons. When this happens, then such person becomes a user of the biometric system.

2.3.2.2 Recognition

Recognition is the biometric function that recognizes persons strictly speaking. However, there are two possible methods to carry out this function: verification and identification.

2.3.2.2.1 Verification (1:1)

Verification is a user recognition process in which the user has to claim his/her identity before the comparison starts. Then the biometric system compares the biometric sample to that claimed user's biometric reference, which has been stored during the enrolment. This kind of comparison is called 1:1. As a result, a similarity score is provided. Depending on this value and the fixed decision threshold, the user is accepted or rejected. The verification is correct either if a user who claims his identity is accepted or a non-enrolled person is rejected. Otherwise, this process will commit an error.

2.3.2.2.2 Identification (1:N)

Identification is a user recognition process in which the biometric sample is compared to all biometric references that have been stored in the data storage subsystem. This kind of comparison is called 1:N. The biometric system returns a candidate list that provided zero, one or several candidates. If a user has been included on the list, the identification is correct. The identification will be wrong either if a user is not included on that list or if a list with one candidate at least is returned for a non-enrolled person.

The identification process can be of two types: open-set identification, in which all kind of people are going to use the biometric system, and closed-set identification, in which only a specific set of people are going to use it.

2.4 Conclusions

This chapter has offered an overview to biometrics with the intention to introduce briefly the science and its related technology in which is encompassed this PhD Thesis. To that end, the fundamental concepts of biometrics and its principles have been described. This description also covers the classification of the existing biometric modalities and the most relevant properties of the different biometric characteristics. Furthermore, this chapter has explained the basis of a general biometric system, including their components and the most significant biometric functions: enrolment and recognition.

Chapter 3

Evaluation of biometric technology

When a new system is developed, one of the most important questions to answer is whether the system works properly or not. That means in what extent the system achieves the objectives for which it was has been designed for. In other words, whether is efficient and fulfils those factors related to accuracy, reliability, security, safety, quality, etc. In order to analyse if a system satisfies these conditions, different kinds of tests have been established.

This chapter is an introduction to the evaluation of biometric technology. First of all, it explains the importance of biometric evaluations and presents a theoretical classification of biometric testing types. As it can be seen in this classification, the most significant type is biometric performance testing. Generally speaking, this kind of test analyses the biometric system operation. This is the most widespread and it is basic for carrying out other types of tests.

As a consequence, the rest of the chapter describes in detail the testing of biometric performance. Initially, a literature review will be provided. This review covers two aspects in particular. On one hand, it examines the evolution of biometric performance testing methodologies, from the first published works to its standardization. On the other hand, this review provides a brief description of the most relevant biometric performance tests already conducted. Then, essential concepts of the standardized biometric performance evaluations as well as the fundamental performance metrics will be explained.

3.1 Biometric evaluation

A biometric evaluation consists of analysing biometric algorithms, components, systems, and/or complete applications to test if they provide specific characteristics or fulfil certain requirements providing empirical evidences [IEEE'09d]. This is a fundamental process that helps developers, customers, integrators and researchers in the following activities:

- To know biometric system behaviour, adjust it and/or improve it,
- To determine advantages and disadvantages of biometric systems,
- To detect operation failures,
- To decide for which applications the biometric system is appropriate, and
- To compare biometric systems and select the best option for a biometric solution.

Opposed to these benefits, there are some inconveniences to consider. A biometric evaluation is expensive [JON'00, WAY'00, MAN'02, ISO'06b, IEEE'09c, DUN'09, LI'09, PET'09]. Although its cost depends on the type of evaluation to conduct, if the evaluation requires real-time biometric data from original users, a significant number of people have to take part in the tests. This fact involves quite a lot of tasks such as recruit users, make that such users interact with biometric devices several times, and take the necessary precautions to guarantee privacy. These circumstances cause that biometric evaluations are time consuming and require a considerable number of resources increasing their price.

However, this drawback is overcome by the benefits obtained. Evaluations should be an indispensable process that all biometric systems shall undergo due to the probabilistic nature of this technology and its dependence of contour conditions as it was explained in Chapter 1. Evaluations are even more important, when the systems are going to be used in applications which entail a high level of security (e.g. national security or ATMs) or in applications which demand a high level of accuracy such as criminal investigations or forensic analysis. In these cases, the existence of a failure may have negatively effects [MAG'11].

On the other hand, the aforementioned characteristics of biometric technology make that biometric evaluations are challenging. Multiple aspects can be checked and each analysis entails to control a wide range of parameters [JAIN'98, WAY'00, MAN'02, ISO'06b, JAI'07, IEEE'09c, DUN'09]. As a result, different types of evaluations exist and all of them require an exhaustive evaluation methodology. The evaluation types have been already identified in literature and the next section describes them.

Nevertheless, this is not the case for the methodologies designed to carry on evaluations. Formal procedures have been mainly specified for only two types of evaluations: biometric performance testing and biometric security testing. It means that there are many aspects that are not tested and may be the cause of system failures [LI'09]. Therefore, the major aim of this Thesis is to specify evaluation methodologies for analysing factors that have been recognized as detrimental to biometric system performance, i.e. environment and user's interaction, but for which an evaluation methodology does not exist. However, the specification of these methodologies is based on the already stated evaluation methodologies. Both provide the

basis for the developed work. For that reason, biometric performance testing will be explained in this chapter whereas biometric security testing will be presented in Chapter 4.

3.2 Types of biometric evaluations

In a similar way to other systems, there are several aspects that can be tested in a biometric evaluation. For this reason, different types of biometric evaluations have been established. However, there is not a consensus between published works [MAN'02, ISO'06b, ISO'07b, IEEE'09d, DUN'09, LI'09]. Some aspects may be analysed measuring similar parameters so, depending on the document the classification of biometric evaluations varies slightly. Furthermore, not all documents list all types.

The next subsections describe the most important types of biometric evaluations based on [IEEE'09d] including other forms of testing that have been defined in the rest of works previously mentioned.

3.2.1 Performance testing

This type of testing consists of measuring biometric system features. Usually, it quantifies the "technical performance" of a biometric system, i.e. its recognition accuracy and processing speed [MAN'02, ISO'06b, IEEE'09d]. This has been the most common biometric performance evaluation in the last three decades [ISO'07b] because performance metrics does not only measure the system's capability to recognize people in terms of accuracy and speed, but also they are used for quantifying the security strength of biometric functions as well as for obtaining usability information related to users that are not able to enrol and verify/identify.

However, there are other kinds of evaluations that involve performance testing. On one hand, there is a group of tests that consists of analysing different biometric system properties apart from accuracy and speed. Such properties can be also considered as part of the biometric system performance. This group includes the following types:

- Reliability testing. This test analyses the frequency of errors as well as the biometric system's ability to continue working when errors occur [IEEE'09d].
- Robustness testing. Robustness tests study the biometric system's ability to operate given noisy data or a low variability [IEEE'09d].
- Availability testing. This type of tests measure the percentage of time that biometric system is able to be used for presenting a biometric characteristic [IEEE'09d].
- Response time testing. These tests quantify the time that a user has to wait for the biometric system decision [IEEE'09d].
- Maintainability testing: this test measures the effort required to maintain biometric system over a short or long term [IEEE'09d].

On the other hand, there are other group of tests that analyse which factors affect biometric system performance and to what extent. These tests entails a kind of technical

performance testing which includes the analysis of influential factors on performance metrics (i.e. error rates and throughput times) as well. Within this category the following types of performance testing may be considered:

- Environmental influence testing. These evaluations study if environmental conditions such as temperature, humidity, illumination, noise, etc, affect to biometric system performance. As it was mentioned in the introduction, the development of an evaluation methodology for this type of performance evaluations is one of the major objectives of this dissertation. Therefore, all related aspects and the achievements obtained in this work have been detailed in Chapter 5.
- Usability testing. Some usability tests measure the influence of different factors related to the user, the biometric system and their interaction on biometric performance [NIST'06a, MOD'06, NIST'06b]. Considering those tests and in a similar way to the preceding evaluation type, another primary objective of this dissertation is the development of an evaluation methodology for conducting them. Likewise, the complete research work carried out on this matter has been fully explained in Chapter 6. However, usability term encompasses a range of issues such as ergonomics, ease of use itself, human factors, user interfaces, user acceptance, and etcetera, which are more focused on users instead of on biometric systems. This type of usability testing will be described separately in section 3.2.5.
- Interoperability testing. These tests determine or compare biometric performance when any kind of interoperability exists between subsystems, signal/image or capture devices [ISO'08, IEEE'09d].
- Scalability testing. Scalability tests analyses the biometric system's ability to adapt itself to a greater size [IEEE'09d]. Typically, the most common test is to quantify performance for one-to-many biometric systems when their database/population has been increased [DUN'09].

3.2.2 Biometric conformance testing

Conformance testing is defined as any activity concerned with determining directly or indirectly that specific requirements are fulfilled [IEEE'09d]. Such requirements are usually pre-established by a standard. In this case, the conformance testing is a process that gives assurance that the product, system or process satisfies those requirements and it is conformant to that standard [LI'09].

In biometrics, there are three relevant conformance tests. These have been established by means of standards which define the technical specification as well as standards which specify the corresponding conformance testing methodology. These tests are the following:

- Conformance testing methodology for biometric data interchange formats. ISO/IEC 19794 [ISO'11a] is a multipart standard which defines interoperable data formats for biometric data of different modalities. In order to analyse the conformity to these

standards, the multipart standard ISO/IEC 29109 [ISO'09c] specifies the evaluation methodology for conducting such tests.

- Conformance testing for the biometric programming interface (BioAPI). The multipart standard ISO/IEC 19784 [ISO'06a] define a common interface that allows the communication between software applications of different biometric technologies. For testing the compliance with these standards, the multipart standard ISO/IEC 24709 [ISO'07d] addressed the necessary testing methodologies.
- Conformance testing for Common Biometric Exchange Formats Framework (CBEFF). The ISO/IEC 19785 [ISO'06c] multipart standard defines a common structure for the exchange of biometric information. In this case, the testing methodologies to assess conformance have been developed by US as the national standard ANSI/ INCITS 473 2011 [ANSI'11].

In addition, the quality assessment of a biometric sample quality can be considered a kind of conformance testing. In this case, the series of standards ISO/IEC 29794 [ISO'09d] defines the quality criteria for some modalities such as finger [ISO'10c], face [ISO'10b] and iris (currently under development)[ISO'12b]. Normally, this test is carried out by the capture device and/or other biometric system component when the biometric sample is acquired and/or processed. The compliance of the standard requirements is often quantified by means of a quality measurement algorithm providing a number called quality score [LI'09].

3.2.3 Security testing

As it was mentioned in Chapter 1, one of the most important uses of biometric technology is for applications that require security. As a consequence, it is necessary to know if biometric systems are secure and the level of security achieved for them. This is the main purpose of a security testing.

Specifically, security testing consists of checking if biometric systems satisfy certain security requirements and studying their resistance to potential attacks. Actually, it is a set of biometric evaluations that involves:

- conformance testing for assessing whether security requirements are fulfilled or not, and
- vulnerability assessment including penetration testing. This part of the security evaluation entails to make a list of potential threats, decide which of them are exploitable and devise specific attacks. Then these attacks shall be executed carrying out the so-called penetration tests. Any successful attack discloses one or more biometric system vulnerabilities.

It is important to note that penetration testing entails biometric performance testing as well. This is because the accuracy of a biometric system quantifies the probability of success of one type of attack called "zero-effort impostor attempt", also considered intrinsic failures of biometric technology [JAI'07, IEEE'09d, LI'09].

3.2.4 Privacy testing

Privacy testing is another kind of conformance testing which analyses whether a biometric system is compliant with privacy regulations or not. Essentially, its purpose is to ensure that personal information (i.e. biometric and personal details) is used appropriately [NSTC'06a]. To that end, privacy testing entails to check if biometric system's implementation provides privacy protections as well as to assess if other related elements such as documentation and administrative procedures fulfil privacy considerations [NSTC'06a, IEEE'09d].

3.2.5 Usability testing

In general, it can be said that usability testing is a type of biometric evaluation that is focused in users and their interaction with the biometric system [LI'09]. Its objective is to quantify till what extent a biometric system can be used. According to ISO 9241 Part 11 international standard [ISO'98], usability is composed by three parameters: effectiveness, efficiency and satisfaction and these are the parameters which are measured in this type of tests.

However, usability is closely related to other issues which have been defined in literature such as:

- Acceptability testing or user acceptance testing, which studies the degree to which people are able to accept the use of a specific biometric characteristic, method or system for biometric recognition [IEEE'09b].
- Ergonomic design which is focuses on the interaction between the user and the biometric system analysing tasks, movements, and user behaviours [LI'09].

In the last years, due to the connection between all these areas, a conceptual model called Human-Biometric Sensor Interaction (HBSI) [ELL'10, KUK'10] has been proposed. The purpose is to explain the relationship between human, biometric capture device and biometric system and study the overall biometric performance considering metrics which come from the overlapping of such areas.

3.2.6 Other kinds of testing

In addition, there are other types of testing that are similar to other technologies. These are the following:

- Cost/benefit testing which involves a trade-off between the cost of biometric system, its operation and maintenance versus its benefits such as performance, security and usability properties as well as the reduction of employment other resources such as human operators, tokens and etcetera [DUN'09, LI'09].
- Personal safety which analyses the potential risk of biometric systems to public health [UKBWG].

3.3 Biometric performance testing

As it was explained in section 3.2.1, biometric performance testing quantifies the technical performance of a biometric system. For achieving such objective, it calculates error rates and throughput rates. Error rates provide biometric system accuracy for enrolling users and recognizing them. Specially, these metrics measure the proportion of users for whom there is a failure during enrolment as well as the proportion of users who have been falsely rejected or accepted by the system. Likewise, throughput rates determine biometric system speed measuring how many persons can process a biometric system per time unit, including the human-system interaction time in addition to the computational processing time of the biometric system.

In general, performance testing is an evaluation which has been and continues to be conducted many times. Whenever researchers and developers would like to know biometric algorithms or systems behaviour, they have to carry out some kind of biometric performance evaluation. However, these evaluations have some inconveniences. Firstly, biometric performance testing is not a straightforward task. Test methods shall establish and control several parameters in order to obtain repeatable and intercomparable results. Biometric performance evaluations are not comparable unless similar requirements and test procedures have been followed. Besides, evaluations performed by vendors are not always reliable due to the fact that they only provide favourable results but not complete information about the testing process. Considering these circumstances, biometric stakeholders have demanded independent evaluations and common evaluation methodologies [JON'00].

The first independent tests can be considered that took place from the late seventies [HAB'76] and the early eighties [RAND'80]. Both evaluations were done for organizations which objectives were to know the capability of speaker and signature verification systems and the authentication capability of keystroke modality respectively. Then, during the 90's different independent performance tests were carried out:

- From 1993 through 1997, the Facial Recognition Technology (FERET) program was performed with the goal of developing algorithms for automatic face recognition. One phase of this program was to assess the proposed algorithms using an independent method and at the same time, to analyse the state of the art in automatic face recognition [JON'96, FERET'11].
- In 1997, 1998 and 1999 the Speaker Recognition Evaluation (SRE) was conducted by the National Institute of Standards and Technology (NIST). Its purpose was to progress in the field of text independent speaker recognition and to measure the performance of this technology [SRE'12]. This evaluation has been also done every year or every two years up to now.
- In 1999, the International Biometric Group (IBG) performed the first round of comparative biometric testing. The aim of these tests was to assess biometric systems performance under controlled, real-world operation conditions [IBG'02, IBG'12].

Nevertheless, each of these evaluations had their own test plan. A normative and general evaluation methodology did not exist. As a result, it was in 1999 when the Biometric Working Group (BWG) decided to specify a generalized methodology for testing and reporting biometric system performance. This document was entitled "Best Practices in Testing and Reporting Performance of Biometric Devices, Issue 1" [BWG'00]. It defined basic metrics and specified minimum requirements for testing and reporting. This document was based on two previous documents written by NIST: "An introduction to evaluating biometric systems" [JON'00] which was also based on the FERET evaluation methodology, and "The NIST Speaker Recognition Evaluation – Overview, methodology, systems, results, perspective" [DOD'00].

This first draft of a biometric performance testing methodology was circulated within the biometric community. It received a lot of comments, especially from the Biometric Consortium WG on Interoperability, Performance and Assurance. Considering these comments and different evaluation reports (i.e. BioIS Study [ZWI'00], Biometric Product Testing Final Report [MAN'01], Facial Recognition Vendor Test Evaluation Report (FRVT) [DUA'01], IBG's comparative biometric testing [IBG'03] and FVC2000: fingerprint verification competition (FVC) [MAI'02]), A. J. Mansfield and J. L. Wayman wrote a second version of this methodology in 2002 [MAN'02]. This version was considered a consistent and comprehensive methodology and it is referenced as the first formal biometric performance testing methodology.

Afterwards, in December of the same year, this methodology was proposed for being an international standard to ISO (International Organization for Standardization) and IEC (International Electrotechnical Commission) organizations. ISO is an independent, non-governmental organization made up of members from the national standards bodies of 164 countries. For the standard development, it is divided in different technical committees (TC) and subcommittees (SC) and each of them is focused on a different subject [ISO'12a]. Likewise, IEC is also a non-profit, non-governmental organization, founded in 1906, which develops International Standards and operates conformity assessment systems in the fields of electrotechnologies. It is made up by national committees of different countries and is divided in technical committees and subcommittees as well [IEC'12].

In 1987, both organizations merged creating the joint technical committee JTC1 for dealing with the Information Technology standardization activities. Within JTC1, the subcommittee SC37 was created in 2002 for addressing biometrics. Moreover, SC37 consists of several working groups (WG) and the evaluation of biometrics is covered by WG5.

As a consequence and after the acceptance of the proposal in March of 2003, WG5 was responsible for the development of the "Biometric performance testing and reporting" project which received the number ISO/IEC 19795 [RON'03]. This project was organized in several parts in order to cover different aspects of biometric performance evaluation.

Finally, in 2006 the first part of this standard was approved as an International Standard (IS) stating the principles and framework for biometric performance testing and reporting [ISO'06b] and, after all, establishing a common and general methodology. During the following

years the rest of the parts were approved, and even further parts have been added. The complete standard as well as its most relevant contents will be described in the next subsections.

Meanwhile, different biometric performance evaluations have been carried out. It could be classified in two groups:

GRAND CHALLENGES

Since 2000 to nowadays, NIST and several institutions and organizations (e.g. the University of Bologna, the University of Surrey, the Hong Kong University of Science and Technology, the Netherlands Forensic Institute and BioSecure Network of Excellence) have conducted a series of independent, open and large-scale challenges in order to know the state-of-the-art and to advance in a specific modality or for checking whether certain technique or data format improves performance for a particular modality. Table 1 shows the most relevant competitions, the institutions that organized those competitions and when these evaluations were conducted for different biometric modalities. Further information could be obtained at the web pages of these institutions [SVC'03, FAC'05, BIO'08, ICDAR'09, BIO'10, FVC'12 and NIST'12].

The philosophy of this kind of challenges has been very similar. It consists of carrying out a technology evaluation (see section 3.5.1). For this purpose, a general wide database is collected which contains the biometric samples that will be used during the evaluation. Likewise, different algorithms that are able to process biometric samples of such database are requested to companies, organizations, academia or any other institution which want to take part in the competition.

In order not to bias evaluation results, the participants do not have access to the overall gallery. They only have access to a training dataset for adjusting their algorithms in a similar way that it would be done in a real application. Biometric samples that belong to the training dataset are excluded later from the gallery in order to avoid negative effects such as overtraining. The objective is that results predict the correct biometric performance. Besides, participants also receive information about how to submit their algorithms and what kind of results will be obtained.

Once all tests were finished, the organizing institution obtains such results and disclose them by means of a report. This report usually describes the complete information about the evaluation, the collected database and shows curves which depict biometric performance per each submitted algorithm.

Table 1. Grand challenges conducted from 2000 until now [SVC'03, FAC'05, FVC'12, NIST'12].

Modality	Title	Organization	Dates
Fingerprint Recognition	Fingerprint Verification Competition (FVC)	University of Bologna	2000, 2002, 2004, 2006 and 2009 (ongoing)
	Fingerprint Vendor Technology (FpVTE)	NIST	2003
	Proprietary Fingerprint Template Evaluations (PFT)	NIST	2003 to 2010 and 2010 (ongoing)
	Minutiae Interoperability Exchange Test (MINEX)	NIST	2004, 2005 (ongoing) and 2007
	Slap Fingerprint Segmentation Evaluation (Slapseg)	NIST	2004 and 2009 (ongoing)
	Fast Tenprint Capture Devices Evaluation	NIST	2007
	Evaluation of Latent Fingerprint Technology (ELFT)	NIST	2007 and 2009
Face Recognition	Face Recognition Technology (FERET)	NIST	1993 to 1997
	Face Recognition Vendor Tests (FRVT)	NIST	2000, 2002 and 2006
	Face Verification Contest (FAC)	University of Surrey	2004
	Face Recognition Grand Challenge (FRGC)	NIST	2005
Iris Recognition	Iris Challenge Evaluation (ICE)	NIST	2005 and 2006
	Iris Challenge Evaluation (IREX)	NIST	2008
Speaker Recognition	Speaker Recognition Evaluation (SRE)	NIST	1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2008 and 2010
Signature Recognition	First International Signature Verification Competition (SVC)	Hong Kong University of Science and Technology	2004
	BioSecure Signature Evaluation Campaign (BMEC2009)	FP6 project- BioSecure Network of Excellence	2009
	ICDAR 2009 Signature Verification Competition	Netherlands Forensic Institute	2009
Multiple Biometrics	Multiple Biometric Grand Challenge (MBGC)	NIST	2007
	BioSecure Multimodal Evaluation Campaign (BMEC2007)	FP6 project- BioSecure Network of Excellence	2007
	Multiple Biometric Evaluation (MBE)	NIST	2009
	Face and Ocular Challenge Series (FOCS)	NIST	2010

COMPARATIVE EVALUATIONS

Moreover, in addition to grand challenges, comparative evaluations have been conducted since 1999. The most significant have been the so-called "Rounds" performed by the International Biometric Group (IBG) [IBG'12]. Each round is a scenario evaluation (see section 3.5.1), in which several biometric systems are tested at the same time. These evaluations analyse biometric systems which are able to capture, store and compare biometric samples in verification mode (i.e. comparison 1:1). The whole test plan is described in [IBG'02], being especially noteworthy the following requirements:

- a controlled indoor environment,
- a test population of 240 non-acclimated test subjects,
- each test subject provides several biometric samples per visit executing enrolment, genuine and impostor attempts, and
- such attempts are performed in two separate visits with six weeks between each visit.

Moreover, further scenario evaluations have been carried out at that time by others institutions such as the National Physical Laboratory (NPL) [MAN'01] and Authenti-Corp [AUT'07] following similar requirements. The complete list of all evaluations from 1999 until now could be seen in Table 2.

Table 2. Comparative evaluations conducted from 1999 until now [MAN'01, AUT'07, IBG'12]

Title	Organization	Dates
Round 1 of Comparative Biometric Testing	International Biometric Group (IBG)	1999
Biometric Product Testing	CESG/BWG Biometric Test Programme National Physical Laboratory (NPL)	2000
Round 2 of Comparative Biometric Testing	International Biometric Group (IBG)	2001
Round 3 of Comparative Biometric Testing	International Biometric Group (IBG)	2001
Round 4 of Comparative Biometric Testing	International Biometric Group (IBG)	2002
Round 5 of Comparative Biometric Testing	International Biometric Group (IBG)	2003
Independent Testing of Iris Recognition Technology (ITIRT)	International Biometric Group (IBG)	2004 to 2005
Round 6 of Comparative Biometric Testing	International Biometric Group (IBG)	2006
Iris Recognition Study (IRS06)	Authenti-Corp	2006
Round 7 of Comparative Biometric Testing	International Biometric Group (IBG)	2009

3.4 ISO/IEC 19795 Biometric performance testing and reporting

Once it has been explained the events from which the ISO/IEC 19795 international standard grew out, this section describes the structure and content of this standard.

ISO/IEC 19795 is a multipart standard which establishes requirements for planning, executing and reporting biometric performance evaluations. Currently, this standard is composed by seven separate documents called "Part". Each part is a standard or a technical report in itself. The parts that constitute the ISO/IEC 19795 are the following:

- ISO/IEC 19795 Part 1: Principles and framework, published in 2006 [ISO'06b].
- ISO/IEC 19795 Part 2: Testing methodologies for technology and scenario evaluation, published in 2007 [ISO'07a].
- ISO/IEC TR 19795 Part 3: Modality-specific testing, published in 2007 [ISO'07c].
- ISO/IEC 19795 Part 4: Interoperability performance testing, published in 2008 [ISO'08].
- ISO/IEC 19795 Part 5: Grading scheme for access control scenario evaluation, published in 2011 [ISO'11c].
- ISO/IEC 19795 Part 6: Testing methodologies for operational evaluation, published in 2012 [ISO'12e].

- ISO/IEC 19795 Part 7: Testing of on-card biometric comparison algorithms, published in 2011 [ISO'11b].

Among all parts, Part 1 is essential because it states the basis of biometric performance evaluations and establishes a common evaluation framework for developing and defining test protocols. Therefore, any evaluation methodology which analyses biometric performance shall be specified under such framework and be in accordance to the defined principles and requirements. This is the reason why Part 1 is of major importance for this Thesis. Next section describes the major contents of this part.

3.5 ISO/IEC 19795 Part 1: Principles and framework

Part 1 is a general introduction to biometric systems and their performance evaluation. Its purpose is literally "… to present the requirements and best scientific practices for conducting technical performance testing" [ISO'06b]. To that end, this document initially explains a general biometric system, its main functions and the types of biometric performance evaluations. Then, it describes requirements for planning a biometric performance evaluation. After that, the document specifies requirements for collecting evaluation data. Such data could be biometric samples, test subjects data, biometric system outcomes or any other data relevant to the evaluation. Next, performance measurements and requirements for analysing the collected data are specified. Finally, the document defines requirements for reporting results.

One of the important concepts stated in 19795-1 are the three different kinds of evaluations (i.e. technological, scenario and operational), which is of major importance to the better understanding of the work in this Thesis. Therefore, to finish this chapter, these types of biometric performance evaluations will be further explained, followed by the description of the fundamental measurements for the performance of biometric systems.

3.5.1 Performance evaluation taxonomy

ISO/IEC 19795-1 defines different types of biometric performance evaluations: technology, scenario and operational evaluation. Basically, it depends on the extent to which the biometric system is assessed, as well as the testing conditions.

3.5.1.1 Technology evaluations

Technology evaluations are designed to analyse biometric performance of one or more biometric algorithms of the same biometric modality using a generic database. Due to the fact that users do no interact with biometric system in real time, this type of evaluations are also called *offline*. These evaluations are considered repeatable as long as test procedures and database are the same. It is important to highlight that biometric performance relies upon requirements and conditions in which the used databases are collected.

This kind of evaluations is able to isolate the user interaction effects from the capacity of biometric algorithms to recognize people itself. Besides, this fact allows executing a large number of comparisons. Due to these advantages, these evaluations are suitable for the first development stages of a biometric system in order to improve and adjust the recognition algorithm.

Furthermore, these evaluations are useful for comparing different algorithms or for analysing the algorithm performance when biometric samples have been collected by different acquisition devices. Besides, these evaluations are appropriate to check the functionality of the subsystems and modules which make up the whole biometric system.

3.5.1.2 Scenario evaluations

Scenario evaluations measure biometric performance of a complete biometric system considering certain conditions which model a specific environment. Such environment is based on a real application and its target population. Biometric samples are processed in real time and every evaluation condition is controlled at all time. Depending of the storage capabilities of the biometric system, these evaluations can be *online*, in which users present their biometric characteristic to the capture device and the result of the recognition attempt is obtained in real time or can be a combination of *online* and *offline*, in which the biometric sample is acquired in real time but further processing is carried out lately. Scenario evaluations are also considered repeatable provided that environment and the test crew are controlled and similar test procedures are applied, although changes in test crew may provide an important drawback for the repeatability of the evaluation.

This type of evaluations analyse the complete biometric system, including the acquisition process, so it is possible to check the influence of environmental conditions or user interaction on the system performance. Likewise, it is possible to measure enrolment and recognition times considering all phases of the process.

Taking into account these characteristics, scenario evaluations are suitable when the target application is known and the target of the evaluation is to be able to predict if any factor is going to affect the biometric system operation. In addition, these evaluations are used for selecting the solution that will be finally adopted, among several commercial products, by analyzing which of them behave better under the conditions of the target application.

3.5.1.3 Operational evaluations

An operational evaluation analyses biometric performance of an overall biometric system when it is working in its real operational environment. Performance results are obtained from the outcome of verification/identification attempts executed by test subjects which are the actual users of the system. This kind of evaluation is comparable to a pilot test. Due to these characteristics, operational evaluations are carried out in real time and the test parameters are measured and recorded but not controlled. As a consequence, these evaluations cannot be repeatable.

One significant difficulty for conducting these evaluations is to know the nature of the test subjects. In other words, to know if the user who is executing the recognition attempt is a genuine user or an impostor user. Nevertheless, there are solutions to solve this problem such as to monitor user's interactions.

This type of evaluations is usually carried out in large-scale projects or projects which entail a long term implementation phase. Specifically, operational evaluations are appropriate for checking whether a biometric system satisfies operational requirements or not, determining if it is necessary to adjust any system parameter or if biometric performance is affected by any operational parameter.

3.5.2 Fundamental biometric performance measures

There are several measurements for quantifying biometric performance. However, ISO/IEC 19795 standard establishes two mandatory kinds of metrics: error rates and throughput rates. In the following subsection, such performance metrics are summarized based on the definitions provided by the standard and considering the different processes and types of biometric system functions.

3.5.2.1 Error rates

Error rates are metrics for quantifying accuracy. These rates measure the number of errors that occur during biometric sample acquisition, its processing, its comparison with the biometric template and the decision processes. These metrics depend on the evaluation effort, i.e. the number of enrolment and recognition attempts and the decision policies.

ERROR RATES RELATED TO ACQUISITION AND SIGNAL PROCESSING STEPS

- Enrolment
    - Failure-to-enrol (FTE) rate: is the proportion of the population for whom the system fails to complete the enrolment process. Enrol failures include failures due to those attempts in which users cannot present his biometric characteristic, or those attempts in which the biometric characteristic cannot be acquired or the from which the biometric reference cannot be generated, either due to restrictions of the biometric algorithm, or by the low quality of the samples acquired.
- Recognition
    - Failure-to-acquire (FTA) rate: is the proportion of the recognition attempts for which the system fails to acquire or localize a biometric sample with enough quality. FTA failures involve attempts in which the biometric characteristic cannot be presented or acquired, attempts for which segmentation or extraction processes fail and attempts in which the biometric sample does not achieve quality thresholds.

Both rates are dependent of the quality criteria as well as the enrolment and acquisition policies respectively. These policies shall be described together with the observed rates.

ERROR RATES FOR COMPARISON AND DECISION PROCESSES

- Biometric system used in verification
  - False Non-Match Rate (FNMR): is the proportion of samples, acquired from genuine attempts, which are falsely declared not to match the biometric reference of the same characteristic from the same user who has provided the biometric sample.
  - False Match Rate (FMR): is the proportion of samples, acquired from zero-effort impostor attempts, which are falsely declared to match the compared non-self biometric reference.

The calculation of these rates is dependent of the information provided by the biometric system. For systems in which it is possible to obtain a similarity score, these rates are a function of the decision threshold (τ). The FNMR rate is the proportion of genuine attempts for which the similarity score is below the matching decision threshold. It can be expressed as

$$FNMR(\tau)= \int_0^{\tau} \Psi_G(s)\, ds \qquad (1)$$

being Ψ_G the genuine probability distribution function. In a similar way, the FMR rate is the proportion of zero effort impostor attempts for which the similarity score is greater than the matching decision threshold. This can be expressed as

$$FMR(\tau)= \int_{\tau}^{\infty} \Psi_I(s)\, ds = 1 - \int_0^{\tau} \Psi_I(s)\, ds \qquad (2)$$

being Ψ_I the impostor probability distribution function [WAY'97]. In case of systems where the outcome is an accept/reject decision, these rates are not a function, but just a single value for the fixed decision threshold.

Furthermore both rates are often depicted together using the Receiver Operating Characteristic (ROC) curve and/or the Detection Error Trade-off (DET) curve. Each point of the curves represents the value of FMR against FNMR (or 1 – FNMR) rates per each decision threshold. The difference between them is the representation of such rates. Typically, ROC curve plots the FMR rate against (1 – FNMR) rate, whereas DET curve plots the FMR rate opposed to the FNMR rate using a normal deviate scale. In addition, both axes can be plotted a linear, semi-logarithmic or logarithmic scale [IEEE'09d, DUN'09].

- Biometric system used for identification

The ISO/IEC 19795 Part 1 standard does not define error rates for identification systems at the attempt level.

ERROR RATES FOR THE COMPLETE RECOGNITION PROCESS

- Biometric system used in verification
    - False Reject Rate (FRR): is the proportion of genuine verification transactions that are incorrectly denied.
    - False Accept Rate (FAR): is the proportion of zero effort impostor transactions that are incorrectly accepted.

These rates include errors due to acquisition failures (related to FTA) or due to matching errors. Depending on the biometric system a transaction will consist of one or more attempts. Therefore both rates are a function of the number of attempts per transaction as well as quality and decision thresholds. In a similar way to FNMR and FMR rates, the relationship between FRR and FAR rates are usually plotted by means of ROC and/or DET curves for a fixed number of attempts per transaction.

Other rates are:

- Generalized False Reject Rate (GFRR): is a general rate for quantifying rejection errors considering the combination of enrolment, acquisition and false non-match errors.
- Generalized False Accept Rate (GFAR): is also a general rate but it quantifies acceptance errors including enrolment, acquisition and false match errors.

GFRR and GFAR are global rates that combine all processes and are useful when comparing several biometric systems. The ISO/IEC 19795-1 standard does not provide a defined method to calculate them. It addresses that such method shall be established in accordance to the evaluation.

- Biometric system used for identification
    - False Negative Identification Rate (FNIR): is the proportion of identification transactions for enrolled users, for whom the correct identifier is not included in the candidate list returned by the system.
    - False Positive Identification Rate (FPIR): is the proportion of identification transactions for non-enrolled users, for whom the candidate list returned by the system is not empty. This rate cannot be only obtained for closed-set identification systems.

These rates are dependent on quality and decision thresholds, the number of identifiers returned at the candidate list, which is called rank. Both rates increase with the database size. Again these rates can be plotted using ROC curves (i.e. representing

FPIR opposed to 1 – FNIR) and DET curves (i.e. representing FPIR opposed to FNIR) for a fixed database size and a fixed rank.

An additional error rate is:

- Identification rate or (True-positive) identification rate: is the proportion of identification transactions carried out by enrolled users for whom the correct identifier is included in the candidate list.

This rate can be expressed as 1 – FNIR, so it is a function of similar parameters to FNIR. Normally, such rate is plotted by means of the Cumulative Match Characteristic (CMC) curve which depicts the identification rate (y-axis) as a function of the rank (x-axis) for a fixed database size.

Finally, the ISO/IEC 19795-1 standard states that performance measurements are affected by two types of errors: systematic errors and random errors. Such errors cause that error rates are subject to an uncertainty which is not quantifiable. As a result, the associated uncertainty of these metrics shall be estimated. In order to do that, the standard proposes some methods for calculating FNMR and FMR uncertainty based on the estimation of the variance and the confidence intervals.

3.5.2.2 Throughput rates

Throughput rates are metrics for quantifying speed. These measurements indicate the number of users that biometric system is able to process per time unit considering both, processing speed and the user interaction. Depending on the system, one or the other factor will be more significant. For example, for biometric systems in verification mode is more relevant the time that user takes to present his biometric characteristic to the acquisition devices than the processing and comparison time. Whereas, for biometric systems in identification mode which have numerous enrolled users, it is more important the processing time needed, as the comparison process takes more time than the user interaction.

For throughput rates, the standard does not establish specific metrics. The document only addresses that it is essential to determine the very instants at which the time will begin and finish considering the biometric system under test. Therefore, both criteria shall be defined and reported prior to tests are conducted.

Analyzing some of the existing biometric performance evaluation reports [MAN'01, IBG'06, IBG'09], normally two relevant measurements are calculated: enrolment and recognition duration time. Besides, the most common way to depict such measurements is by means of metrics that provide information about the entire set of measured values such as the following:

- The arithmetic mean (μ) of the duration time which describes the central tendency of the collection of measured times. This is obtained using the following equation:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} t_i \quad (3)$$

being "n" the total number of measurements of the duration time and "t" the measured time per each users' enrolment/recognition attempt or transaction.

- The standard deviation (σ) of the duration time which shows the variation of the collection of measured times from the mean. This is obtained using the following equation:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_i - \mu)^2} \quad (4)$$

being "μ" the arithmetic mean, "n" the total number of measurements of the duration time and "t" the measured time per each users' enrolment/recognition attempt or transaction.

- The minimum time, which is the smallest time of all measured times.
- The maximum time, which is the highest time of all measured times.

These metrics are usually expressed in seconds. Together with these metrics, it is essential to report some information about the characteristics of the processing system used to obtain those times. These data are important because the time is hardware dependent. Therefore, system details such as the specific type of processor, its speed, its memory capacity, the OS installed and the particular program interface used to obtain those times shall be given.

3.6 Conclusions

This chapter has described the state-of-the-art of the evaluation of biometric technology. The first major objective of this PhD Thesis is to evolve in the area of biometric system testing developing evaluation methodologies, in order to quantify the effects of contour conditions on biometric performance. Therefore, this overview was necessary to help the reader to understand the starting point of the work in this PhD Thesis.

As a consequence, this chapter have provided an introduction to the evaluation of biometric technology. Firstly, the importance of biometric testing has been described, followed by a classification of the biometric evaluation types. This classification allows explaining which kinds of biometric evaluations exist and the current needs in this field.

After that, the biometric performance testing has been detailed. This is the most significant evaluation type and set the grounds for the research work carried out in this PhD Thesis. Consequently, this chapter has described the evaluation methodologies from the first versions to its standardization as the ISO/IEC 19795 standard. Moreover, the most relevant contents of this standard as well as the biometric performance tests that have been already conducted have been explained.

Chapter 4

Security evaluation of biometrics

One of the most common applications of biometrics is within the information technology (IT) security field. Biometric technology is used in security systems instead of, or in addition to, passwords and/or tokens, due to its properties for people identification. Biometric functions provide proper security mechanisms for protecting information against unauthorized users. Considering this fact and the lack of evaluation methodologies in biometrics, the security evaluation of biometric systems has been traditionally performed according to IT security evaluation methodologies.

This chapter explains the evolution of IT security evaluation methodologies and their development till the current common evaluation framework "Common Criteria for Information Technology Security Evaluation (CC)". Then, an overview of this standard will be provided including the description of its parts as well as their contents. In addition, an explanation of the major security concepts, the terminology used and the Common Criteria general evaluation model will be given.

Finally, the chapter is focused on biometrics and the difficulty to apply Common Criteria evaluation methodology to this technology. Security statements and previous guidelines developed for testing biometric products in the context of Common Criteria will be detailed. Furthermore, those evaluations of biometric systems that have already been conducted using Common Criteria will be mentioned.

4.1 Information Technology security evaluations

In the field of Information Technology, security evaluations have been carried out since the last decades of the 20th century. At the beginning these evaluations were conducted by each country according their own methodologies and standards, being the most relevant the following:

- The DoD 5200.28-STD titled "Department of Defense Trusted Computer System Evaluation Criteria" (TCSEC) also known as "The Orange Book" [TCSEC'85] that was applied in United States of America. It consisted of a set of technical security criteria and the corresponding technical evaluation methodologies developed for supporting the DOD Directive 5200.28 "Security Requirements for Automatic Data Processing (ADP) Systems". This DoD directive was published in 1972 and the complementary manual DOD 5200.28-M [DoD'73] was published a year later. Then, several researches and works were developed by different organizations till 1983 when the CSC-STD-00l-83 version [TCSEC'83] was published. Such version was updated in 1985. Finally, this standard was cancelled in 2002 by the DoD Directive 8500.1.
- The "Information Technology Security Evaluation Criteria" (ITSEC) [ITSEC'91] which was used in Europe. These criteria were developed by France, Germany, Netherlands and United Kingdom in 1991 with the objective to harmonize security criteria of different European countries. This work was based on the US Orange Book as well as on works that already existed such as the British CESG Memorandum Number 3 [CESG3'89] and DTI Commercial Computer Security Centre Evaluation Levels Manual [DTIEC'89], the German Criteria for the Evaluation of Trustworthiness of Information Technology Systems [ZSIEC'89] and the French Catalogue de Critères Destinés à évaluer le Degré de Confiance des Systèmes d'Information [SCSSI'89].
- The "Canadian Trusted Computer Product Evaluation Criteria" (CTCPEC) [CTCPEC'93] which was published in 1993 by the Canadian System Security Center. These criteria were developed with the intention to update the US Orange Book according to the IT products evolution at that time. Due to this standard being based on the US Orange Book and considered some of the British and German documents used during the development of ITSEC, it can be said that it was a preliminary attempt to combine TCSEC and ITSEC criteria.

Due to these initial works, an official project was started to join American and European IT security criteria and to replace the TCSEC criteria for the US government side. NIST and the National Security Agency (NSA) developed the "Federal Criteria for Information Technology Security" (FC). Two drafts [FCITS'92b, FCITS'92a] were released for public review and comments in December of 1992. The idea was that both documents became new Federal Information Processing Standards (FIPS). Nevertheless, neither of them achieved that stage. The Common Criteria international project begun and the Federal Criteria documents were abandoned. However, some ideas of Federal Criteria project were retained in Common Criteria.

The Common Criteria was initially a project of the ISO organization. In 1990, this organization began the development of a new standard that would align all the existing IT security evaluation criteria and would allow the mutual recognition of IT security evaluations. That standardization activity was assigned to the ISO/IEC JTC1 SC27 WG3. However, this work did not progress very fast because it entailed a significant amount of work and negotiations between different nations [CC-1'99].

Therefore, it was in 1993 when the sponsoring organizations of the TCSEC, ITSEC, CTCPEC and FC decided to form a group that works in parallel for supporting ISO efforts. Such group was called "Common Criteria for Information Technology Security Evaluation Project" (CC), normally known as Common Criteria (CC). As a consequence, experts from United States, Canadian and European nations (France, Germany, United Kingdom and Netherlands) began the development of CC. Obviously, this work was based on all the existing IT security evaluation criteria above mentioned: TCSEC, ITSEC, CTCPEC and FC.

As a result, three drafts were prepared in 1994. These documents were circulated for review and comments and finally, in 1996 the first version of Common Criteria was distributed, which consisted of a set of four documents [CC-1'96, CC-2'96, CC-3'96, CC-4'96]. From that initial version, more nations became involved in this project and several revisions and new versions have been done. Some of these works have been published as CC standard, others have been published as both CC and ISO/IEC standards and a few of them have only achieved a draft stage. The complete list can be seen in Table 3 being version 3.1 Release 4 [CC-1'12, CC-2'12, CC-3'12, CEM'12] the current approved version.

Table 3. List of the different versions of CC and ISO/IEC standards

Version	Title	Year
1.0 (trial version)	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements o Part 4: Predefined Protection Profiles	1996
2.0	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements	1998
2.0 (with minor modifications)	• ISO/IEC 15408, Information technology – Security techniques – Evaluation criteria for IT security o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements	1999
CC 2.1 and CEM 1.0	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements • Common Methodology for Information Technology Security Evaluation	1999
CC 2.2 and CEM 1.2	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements • Common Methodology for Information Technology Security Evaluation	2004

2.3	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements • Common Methodology for Information Technology Security Evaluation	2005
Version 2.3	• ISO/IEC 15408 – 1, Information technology – Security techniques – Evaluation criteria for IT security o Part 1: Introduction and general model o Part 2: Security functional components o Part 3: Security assurance components • ISO/IEC 18045, Information technology – Security techniques – Methodology for IT security evaluation	2005
3.0 (draft version)	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements • Common Methodology for Information Technology Security Evaluation	2005
3.1	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements • Common Methodology for Information Technology Security Evaluation	2006
3.1 Release 2	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements • Common Methodology for Information Technology Security Evaluation	2007
3.1	• ISO/IEC 15408 – 2, Information technology – Security techniques – Evaluation criteria for IT security o Part 1: Introduction and general model o Part 2: Security functional components o Part 3: Security assurance components • ISO/IEC 18045, Information technology – Security techniques – Methodology for IT security evaluation	2008 (except Part 1 in 2009)
3.1 Release 3	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements • Common Methodology for Information Technology Security Evaluation	2009
3.1 Release 4	• Common Criteria for Information Technology Security Evaluation o Part 1: Introduction and general model o Part 2: Security functional requirements o Part 3: Security assurance requirements • Common Methodology for Information Technology Security Evaluation	2012

4.2 Common Criteria for IT security evaluation

As mentioned above, Common Criteria for Information Technology Security Evaluation [CC] is an international standard evaluation framework for carrying out security evaluations of IT products. Briefly, this framework states a set of functional requirements for describing the security functionality of an IT product as well as a set of assurance requirements to fulfil when such product is assessed. Furthermore, this framework specifies the evaluation methodology to apply in compliance with the assurance requirements.

To address such evaluation framework, Common Criteria is a multipart standard which is made up by three parts. In addition, there is a companion document to CC which is the Common Evaluation Methodology (CEM) standard. All of them are described below.

- Part 1: Introduction and general model [CC-1'12]. This is an introductory document that explains the principles of IT security evaluations and establishes the Common Criteria evaluation model. Moreover, this part of the standard provides a description of the rest of the parts and explains the essential Common Criteria concepts.
- Part 2: Security functional requirements [CC-2'12]. This part of the standard establishes a set of security functional components for describing in a standard language the security functionality requirements that must meet the IT product under evaluation.
- Part 3: Security assurance requirements [CC-3'12]. In the same way to Part 2, this part establishes a set of security assurance components for describing in a standard language the security assurance requirements that shall be satisfied during a security evaluation. These requirements are usually organized in packages called Evaluation Assurance Levels (EALs) so this part also establishes these levels and the mandatory components that compose each level.
- Common Methodology for Information Technology Security Evaluation [CEM'12]. This is a standard that states the specific methodology for conducting Common Criteria evaluations. It is not a part of Common Criteria standard but has a close relationship with Part 3. More precisely, the document establishes the minimum actions that evaluators shall carry out for applying the assurance measures addressed in the assurance components of Part 3. It is important to note that CEM does not cover all assurance components. This document only provides guidance for those components for which a consensus about the evaluation procedures to perform has been already reached.

4.2.1 Key concepts of Common Criteria

Once the structure of CC standard has been described, this section explains the evaluation model used for this type of evaluations. Due to some concepts are characteristic of Common Criteria, a basic glossary is presented before.

4.2.1.1 Basic glossary

This small glossary defines the most important concepts that are indispensable to understand the Common Criteria. The specific definition are similar to the definition provided in CC Part 1[CC-1'12].

Target Of Evaluation (TOE)

Def.: Set of software, firmware and/or hardware possibly accompanied by guidance.

There are different types of IT products and the same IT product may have different configurations. The TOE is the IT product, or the part of an IT product, or the set of IT products that is going to be assessed considering only the selected configuration/s to be tested.

TOE Security Functionality (TSF)

Def.: Combined functionality of all hardware, software, and firmware of a TOE that must be relied upon for the correct enforcement of the security functional requirements.

In other words, the TSF is only the part of the TOE that provides its security functionality.

TSF Interfaces (TSFI)

Def.: Means by which external entities (or subjects in the TOE but outside of the TSF) supply data to the TSF, receive data from the TSF and invoke services from the TSF.

The TSF interfaces allow to access to the TSF resources as well as to interchange information between the TSF and the external entities. These interfaces also establish the boundaries of the TSF.

Security Problem Definition (SPD)

Def.: Statement which in a formal manner defines the nature and the scope of the security that the TOE is intended to address.

This statement consists of a combination of: threats to be countered by the TOE and its operational environment, the organizational security policies enforced by the TOE and its operational environment, and the assumptions that are upheld for the operational environment of the TOE.

Security objectives

Def.: Statement of the intent to counter identified threats and/or satisfy identified organisation security policies and/or assumptions.

In order to solve the SPD, CC establishes two kinds of security objectives: security objectives for the TOE and security objectives for the operational environment.

Security Functional Requirements (SFRs)

Def.: A translation of the security objectives for the TOE into a standardised language.

CC organizes the SFRs hierarchically in classes, families and components according to the topic of the security objectives that the SFRs expect to satisfy. The current version of the standard specifies in its Part 2 the following classes:

- Class FAU: Security audit
- Class FCO: Communication
- Class FCS: Cryptographic support
- Class FDP: User Data Protection
- Class FIA: Identification and authentication

- Class FMT: Security Management
- Class FPR: Privacy
- Class FPT: Protection of the TSF
- Class FRU: Resource Utilisation
- Class FTA: TOE Access
- Class FTP: Trusted Path/Channels

Security Assurance Requirements (SARs)

Def.: A description of how assurance is to be gained that the TOE meets the SFRs.

In a similar way to SFRs, CC organizes also the SARs hierarchically in classes, families and components but, in this case, according to the intention of the assurance requirement. Likewise, the current version of the standard specifies in its Part 3 the following classes:

- Class ACO: Composition class.
- Class ADV: Development class.
- Class AGD: Guidance documents class.
- Class ALC: Life-Cycle support class.
- Class APE: Protection Profile Evaluation class.
- Class ATE: Tests class.
- Class ASE: Security target evaluation class.
- Class AVA: Vulnerability assessment class.

Evaluation Assurance Levels (EALs)

Def.: Set of assurance requirements drawn from CC Part 3, representing a point on the CC predefined assurance scale that forms an assurance package.

Currently, Part 3 establishes seven EALs. These levels are ordered from lowest (EAL1) to highest (EAL7) considering an increase of the level of assurance to apply during the evaluation. That increase entails an increase of the scope, depth and rigour of the tests and as a consequence, the level of effort devoted to the evaluation is higher. Such EALs are the following:

- Evaluation assurance level 1 (EAL1): functionally tested.
- Evaluation assurance level 2 (EAL2): structurally tested.
- Evaluation assurance level 3 (EAL3): methodically tested and checked.
- Evaluation assurance level 4 (EAL4): methodically designed, tested, and reviewed.
- Evaluation assurance level 5 (EAL5): semi-formally designed and tested.
- Evaluation assurance level 6 (EAL6): semi-formally verified design and tested.
- Evaluation assurance level 7 (EAL7): formally verified design and tested.

Protection Profile (PP)

Def.: Implementation-independent statement of security needs for a TOE type.

A PP is a document that specifies all details about a type of IT products and its proper evaluation. Mainly, the PP explains the characteristics of the TOE type, the security problem definition for it, the security objectives to solve such security problem, the minimum functional requirements that shall meet that TOE type to fulfil those security objectives and the minimum assurance requirements for testing it. Its purpose is to help customers and developers to define CC evaluations for a group of IT products with similar characteristics.

Security Target (ST)
Def.: Implementation-dependent statement of security needs for a specific identified TOE.

The ST is similar to a PP but it is particularized for the specific TOE to assess. Basically, the ST explains the characteristics of the TOE, its security problem, the security objectives to solve such security problem, the functional requirements that shall meet it to comply with the defined security objectives and the assurance requirements for testing that TOE. It is important to emphasize that a ST may be based on (or claim conformance to) a PP or not.

4.2.1.2 Common Criteria evaluation model

CC evaluations are based on the idea that every IT product that protects certain assets may be subjected to attacks. In order to remove, diminish or mitigate such potential threats, it is necessary to impose different countermeasures. CC distinguishes two kinds of countermeasures: IT countermeasures which are provided for the IT product and non-IT countermeasures which are provided by the operational environment. A CC evaluation has the objective to demonstrate that those countermeasures are sufficient to protect the assets as well as those countermeasures are correctly implemented by the IT product (or TOE) and it does not have vulnerabilities. However a CC evaluation only analyses the IT countermeasures, i.e. CC only analyses the TOE whereas it assumes that the non-IT countermeasures are properly addressed by the operational environment.

The aforementioned process is carried out from the analysis of the ST in order to check that countermeasures are sufficient. The ST is a security statement that describes the TOE, the assets to protect and the security problem (SPD). Specifically, the SPD consists of the description of: threats to be faced by the TOE, policies required to the TOE and assumptions of its operational environment. From that SPD, the ST also defines the security objectives for the TOE and the security objectives for the operational environment. As it was previously mentioned, for the first group of objectives, CC specifies in its Part 2 a collection of Security Functional Requirements (SFRs). Using these requirements, the ST expresses the security objectives to be met by the TOE. The second group of objectives are not evaluated. The ST has to describe them but it is assumed that the operational environment of the TOE complies with them.

In addition, a CC evaluation verifies that countermeasures are correctly implemented by the TOE. For doing that, the TOE is analysed for checking that it behaves in accordance with the ST specification. Also, it is checked that the TOE does not have exploitable vulnerabilities

carrying out penetration tests. During this process, not only the SFRs compliance is tested but also other issues related to the TOE such as the development environment, the TOE documentation, etc.

To conduct both types of analysis, CC defines in its Part 3 a collection of security assurance requirements (SARs) and a set of assurance packages named evaluation assurance levels (EALs) as it was already explained. The selected EAL for the evaluation and the corresponding SARs are also specified in the ST. Besides, CC also provides the CEM document which specifies the particular evaluation activities to be carried out per each assurance component. The intention of this document is that CC evaluations are carried out following the same methodology as a guarantee for obtaining objective and repeatable results.

Apart from ST/TOE evaluations, CC evaluation model also covers the evaluation of Protection Profiles (PP). A PP is a security statement similar to an ST but for a TOE type in which essential security requirements for a group of IT products are defined. Due to this fact its evaluation is more or less the same as the ST analysis. It consists of checking that the PP is complete, consistent as well as the proposed countermeasures solve the SPD for the particular TOE type. Nevertheless, CC establishes specific SARs and its associated testing methodology for PP evaluations.

4.3 Common Criteria & Biometrics

After describing Common Criteria, it is important to mention that this evaluation framework has certain limitations. This framework has been specified to address the evaluation of a wide range of IT products and their different technologies. For this reason, most criteria are general and for some technologies it is necessary additional guidance in order to interpret CC and CEM appropriately. This is the case of biometrics [BTSE'01, BEM'02, DUN'09, LI'09, ISO'09a].

The first document developed to deal with biometric evaluations in the context of CC was titled "Biometric Technology Security Evaluation under Common Criteria" (BTSE) [BTSE'01] which was written by Electronic Warfare Associates-Canadian Ltd in 2001. One year later, the Biometric Evaluation Methodology Working Group (BEM WG) produced the Biometric Evaluation Methodology Supplement (BEM) [BEM'02] with the intention to supplement the Common Methodology during the evaluation of biometric systems. However, none of these documents was totally accepted by Common Criteria community and in few years they became outdated due to the publication of new versions of CC. Moreover, both refer to preliminary performance evaluation methodologies that were implemented before the development of current standards.

Years later, in 2009, ISO published the international standard ISO/IEC 19792 Information technology - Security techniques – Security evaluation of biometrics [ISO'09a] which was developed by ISO/IEC JTC1 SC27, i.e. the same group that developed the ISO/IEC versions of the CC standards. Such standard was created with the same objective to address security

evaluations of systems which use biometric technology in compliance with CC. It defines the major requirements to follow, but it neither states a concrete methodology, nor establishes a correspondence between those requirements and the testing activities addressed in CEM.

Furthermore, a new document titled Characterizing Attacks to Fingerprint Verification Mechanisms [CCN'08] has being developed since 2008. This document addresses the vulnerability analysis of biometric systems based on fingerprint modality. Currently, this work continues being discussing into the CC community and comments have been requested to national schemes. Nevertheless, this contribution only addresses a portion of CC evaluations and it only focuses in one modality.

In spite of those attempts to fill the gap between Common Criteria and biometrics and the importance of CC as the only one current formal methodology for security testing of biometric systems, very few CC evaluations have been performed for this technology. Regarding ST/TOE evaluations only five biometric-based products have been certified (as it is shown in Table 4) from more than 1600 certificates issued according to the CC website [CC]. It means a percentage less than 0.31% which is very low in comparison with other authentication technologies such as smart cards (the number of certified products for IC's, smart cards and related devices and systems is currently 508, i.e. 31.7% from the total amount of certificates).

Furthermore, not all of these evaluations include the assessment of the biometric recognition capability of the TOE. It was a must in the old versions of CC in which the vulnerability assessment required the calculation of the strength of the functions (SoF). However, for the current CC version it is not a requirement. For example, for the last ST/TOE evaluation, i.e. the evaluation of the Authentest Server v1.2.6, the biometric system performance was not tested. Its ST [ST'10] said that error rates were already tested by an independent laboratory and consequently, its evaluation is out of the scope of the CC evaluation. Therefore, it cannot be considered a CC evaluation of a biometric ST/TOE.

Table 4. List of ST/TOE evaluations in the field of biometrics

Title	Organization	Year	Compliance
BioscryptTM Inc. Enterprise for NT Logon Version 2.1.3	L-1 Identity Solutions, Inc	2001	CC Version 2.1 EAL2
KnoWho Authentication Server v1.2.2 and Private ID v2.1.15	Iridian Technologies, Inc.	2003	CC Version 2.1 EAL2
VoiceIdent Unit 1.0	Deutsche Telekom AG / T-COM	2007	CC Version 2.3 EAL2 +
PalmSecure SDK Version 24 Premium	Toshimitsu Kurosawa Fujitsu	2008	CC Version 3.1 R2 EAL2
Authentest Server v1.2.6	Authenware	2010	CC Version 3.1 R3 EAL2 +

Considering PP evaluations, the same situation occurs. Table 5 shows all the certified PPs since CC appeared. Nowadays just three PPs are still available from a total of 132.

Table 5. List of PP evaluations in the field of biometrics

Title	Organization	Year	Compliance
Biometric Device Protection Profile (Draft)	UK Government Biometrics Working Group	2001	CC Version 2.1 EAL4
U.S Government Biometric Verification Mode Protection Profile for Medium Robustness Environments	Information Assurance Directorate	2003	CC version 2.1 Obsolete PP EAL4
Biometric Verification Mechanisms	Marcus Krechel, Nils Tekampe TÜV Informationstechnik GmbH	2005	CC version 2.1 EAL2 +
U.S Government Biometric Verification Mode Protection Profile for Basic Robustness Environments	Information Assurance Directorate	2006	CC version 2.1 Obsolete PP EAL2
U.S Government Biometric Verification Mode Protection Profile for Basic Robustness Environments	Information Assurance Directorate	2007	CC version 3.1 R1 Obsolete PP EAL2 +
U.S Government Biometric Verification Mode Protection Profile for Medium Robustness Environments	Information Assurance Directorate	2007	CC version 3.1 R1 Obsolete PP EAL4 +
Biometric Verification Mechanisms Protection Profile	Nils Tekampe, Boris Leidner TÜV Informationstechnik GmbH	2008	CC version 3.1 R2 EAL2
Fingerprint Spoof Detection Protection Profile Based on Organisational Security Policies	Boris Leidner, Nils Tekampe TÜV Informationstechnik GmbH	2010	CC version 3.1 R3 Not EAL

Due to these circumstances and as it was explained in the introduction, this dissertation is also focused on the improvement of previous documents with the aim that biometric systems be certified following the CC certification scheme in a similar way than the rest of IT products. The works developed on this matter have been described in depth in Chapter 7.

4.4 Conclusions

This chapter has offered a review of biometric evaluations considering the security perspective. This is important as the second major objective of this Thesis is the formalization of the developed evaluation methodologies, according to the only security evaluation framework that currently exists, i.e. Common Criteria.

For this reason this chapter has presented an overview of such evaluation framework, explaining the content of CC documents and the evaluation model that this standard proposed. Moreover, a review of the state-of-the-art related to the biometric security evaluations conducted in the context of CC has been provided. This revision shows that the difficulties to apply CC to biometric technology are still unsolved. Therefore, there is a need for further work in that direction.

Chapter 5

Evaluation methodology for environmental testing of biometric systems

Environment is one of the most important aspects that influences performance in biometric systems. Both biometric characteristic and biometric capture device are involved in the acquisition of the biometric sample and can be adversely affected by environmental conditions. As a result, samples may not be acquired or their quality may not be good enough.

Currently, no methodology exists to evaluate such influence, so this dissertation provides a contribution to eliminate such gap in the evaluation of biometric systems. Therefore this chapter establishes an evaluation methodology for analysing the influence of environment on biometric systems performance. This methodology is based on the existing ISO/IEC 19795 multipart standard that addresses biometric performance testing and considers requirements and practices followed in standards that cover the same type of evaluations in other areas.

Firstly, the chapter describes the proposed methodology including its principles, the specification of environmental conditions that should be analysed, the appropriate test procedures and the most significant metrics and measurements to quantify both, biometric systems performance variations and the particular environmental conditions that cause them. After that, the experimental evaluations carried out to develop and validate the proposed evaluation methodology and their results will be shown.

5.1 Overview

Environment is defined as the surroundings or conditions in which a person, animal, or plant lives or operates [OXF'10]. After some years evaluating biometrics technology, researchers and customers have realized that biometric products used in different applications do not behave in the same way as the results obtained in a performance evaluation carried out previously. Therefore it has been noted that environment is a factor which can modify biometric systems performance.

A. Jain, R. Bolle and S. Pankanti [JAIN'98] described the dependence of technology performance on the type of application. They pointed out that the application environment influences directly in the repeatability and distinctiveness of the biometric measure. For this reason they specified seven application categories: cooperative vs. non-cooperative, overt vs. covert, habituated vs. non-habituated, attended vs. non attended, standard environment vs. non-standard environment, public vs. private and open vs. closed. In addition, they explained that test results are dependent upon the specific "real-world" application. Lately, this statement was corroborated in other works such as A.J. Mansfield and J.L. Wayman [MAN'02] and J. Wayman, A. Jain, D. Maltoni and D. Maio [WAY'04]. The former states that performance curves are very application, environment and population dependent. Moreover, it contains an annex which details environmental factors and the corresponding affected biometric modality. The latter explains that changes in the application environment cause a significant impact on the biometric devices performance and also specifies a similar classification of the biometric applications than [JAIN'98]. Most recently books also refer to this problem. T. Dunstone and N. Yager [DUN'09] explain that one factor that affects biometric sample quality is environment and Stan Li and A. Jain [LI'09] mention environment as a source of biometric sample variability.

Likewise, many studies about different biometric modalities claimed the influence of environment in the capability of biometric capture devices to acquire biometric samples (e.g. [DUN'09] and [SAN'09]), in the quality of the acquired samples (e.g. [KIM'03], [KANG'03] and [PRO'11]) or in the overall biometric systems performance (e.g. [KUK'04], [SAN'09] and [BEV'10]).

In view of these works, environment must be considered as a relevant factor that can affect biometric performance negatively. Specifically, its influences in the two main components involved in the first part of the recognition process: the biometric characteristic by itself, and the biometric capture device. Together, these components are responsible for the acquisition of the biometric sample. If one of them or both do not work properly, biometric samples cannot be acquired or the quality of the biometric samples can be insufficient. In both cases, biometric systems performance is reduced and, as a consequence, the level of security of the corresponding application may not be assured. Therefore, it is essential to quantify the influence of environment in biometric system performance.

This chapter describes an evaluation methodology for carrying out the environmental testing of biometric system performance, as well as experimental results obtained after its application. Specifically, the next section explains the concept of environmental testing of

biometric systems performance evaluations. This includes the definition of this kind of evaluations, its principles and scope. Then, the forth following sections specify protocols and requirements that composed this methodology. In particular, section 5.3 covers environmental conditions, its establishment, maintenance and measurement. Section 5.4 states the test plan for biometric systems performance evaluation considering the previous environmental conditions specification. In addition, this test plan explanation is focused on those procedures that are different from a traditional biometric performance evaluation, due to the analysis of the environmental parameters. Section 5.5 determines test execution according to the test plan and section 5.6 describes the calculation of test results and reporting requirements. After that, the following section shows the experimental results of the evaluations accomplished to validate the proposed methodology.

5.2 Environmental testing of biometric systems

An environmental test is defined such as a test conducted under specific environmental conditions, to determine whether these conditions affect the performance, safety or integrity of the materiel or the physical system ([DEF'06], [MIL'08] and [McG'11]). Depending on the source, these tests are specified for natural or simulated environments. Typically, this kind of tests is focused on determining the quality and/or useful life of products. Such aspects are quantified by detecting that systems, materiel or components have not suffered any damage or checking that no mechanical, electrical or chemical failures exist. For doing that, standards determine to use suitable methods such as: visual examination, functional tests, x-ray/radiography, several physical measurements (mass, dimensional measurements, density, etc.), physical tests, non-destructive tests, etc.

As mentioned above, there is a need to carry on environmental tests for biometric systems. However, in case of biometrics, the above mentioned methods are not appropriate. The proper method would be a kind of functional test in which a group of users interact with the biometric system with the objective to calculate the accuracy and speed of the recognition algorithms, and this shall be carried out when both, users and system, are exposed to different environmental parameters. Therefore, this kind of evaluation can be considered as an "end-to-end" biometric system performance evaluation, which is conducted in specific environmental conditions.

As it was explained in Chapter 3, ISO/IEC 19795-1 establishes two kinds of "end -to-end" biometric performance evaluations: scenario and operational evaluations. Both types would be suitable for environmental testing. However, the proposed environmental testing methodology has been described only for scenario evaluations. The reason is because of the fact that one major objective of this dissertation is to merge this methodology with CC and CEM. CC and CEM claims objectivity and repeatability and in that sense the biometric performance evaluations that are characterized by being conducted with a careful control of evaluation conditions are scenario evaluations.

In addition, it is necessary to clarify the concept of environment in order to focus which aspects are covered by the proposed environmental tests. For biometric systems, environment can be understood such as:

- the specific physical location of biometric system including different equipments and apparatus connected or not to the biometric system and which are allocated in its surroundings,
- the personnel that interact with the biometric system, as well as
- the ambient conditions, i.e. all atmospheric parameters (e.g. temperature, humidity, atmospheric pressure, etc.) and other physical and chemical phenomena (e.g. illumination, noise, vibration, mist, dust, etc.) that are presented where the biometric system is to be located and to be used during its operation.

In spite of all these aspects being considered into the proposed methodology, the environmental testing addressed in this chapter only analyses the influence of ambient conditions. Moreover, it is important to emphasize that ambient conditions can affect several elements involved in the recognition process (i.e. test subject biometric characteristic and biometric system or the capture device). However, the proposed evaluation methodology does not distinguish which of them is affected. Its purpose is to quantify the overall influence obtained in the biometric system performance. Due to these circumstances it also indispensable to specify that the proposed methodology only entails *online* testing. *Offline* testing is not suitable in this case because this type of testing is not able to analyse all possible influential effects.

5.2.1 Basic concepts for environmental testing of biometric systems

Before the description of the evaluation model that has been established for the environmental testing methodology of biometric systems, fundamental concepts must be explained.

Environmental conditions

Def.: all atmospheric parameters and other physical and chemical phenomena that can surround the biometric system and influence on its performance.

As it has been mentioned, the term "environmental conditions" entails more aspects than "ambient conditions". However, in the standards that address environmental testing for other kind of systems [DEF'06, MIL'08], the term "ambient conditions" refers to conditions that occurs naturally in contrast to conditions that have been induced. Therefore and considering that the proposed evaluation methodology is focused only on ambient conditions, it has been preferred to use the term "environmental conditions".

Also, it is important to distinguish two concepts related to this term:

- Operational environment: the environmental conditions under which the biometric system is expected to operate. This concept does not associate any predefined value.
- Extreme conditions: environmental conditions that entail very high or very low values and may be hostile for systems operation or even human life.

Evaluation configuration

Def.: physical layout of the environment in which the biometric system is going to be tested including the necessary equipments for performing tests.

Within environmental testing there are two typical kinds of equipments:

- Environment generator: equipment used to establish and maintain the controlled conditions of the test (e.g. an air conditioning system).
- Instrument: calibrated equipment used to measure and/or record environmental parameters (e.g. a thermometer).

Sometimes, the test equipment may include both an environment generator and a measuring instrument (e.g. a climatic chamber).

Evaluation Environment

Def.: environment in which the biometric system is evaluated considering the environmental conditions and the evaluation configuration.

There are two types of evaluation environments:

- Reference evaluation environment (REE). This is the evaluation environment in which the biometric system is analysed to obtain baseline performance metrics for making comparisons.
- Target evaluation environment (TEE). This evaluation environment in which the biometric system is analysed to obtain performance metrics for studying the influence of certain environmental conditions, by comparing with the results obtained at the REE.

Evaluation conditions

Def.: each of the evaluations carried out in a different evaluation environment to assess the performance of the biometric system in one or more specific environmental conditions.

Parties involved in the evaluation

Def.: entities or organizations which are interested in the evaluation and have responsibilities in the evaluation process.

These entities are basically two: the test laboratory which is going to conduct the evaluation and the developer or customer who request to carry out the evaluation. In case the developer is different from the customer (e.g. an end-user requesting to know the

performance of a commercial product), a third entity is added to the number of parties. Test subjects are not considered a party of the evaluation although they have to take part in it.

5.2.2 Evaluation model for environmental testing of biometric systems

Environmental testing entails to conduct two (or more) scenario evaluations: one in the REE and another (or others) in the TEE(s). The evaluation environments will be identical, including the same test subjects, following the same procedures, except for the environmental conditions. The environmental conditions are specific of each evaluation environment. Every evaluation environment is characterized by a set of environmental parameters to analyse and fixed values for such parameters which are named evaluation conditions.

During the scenario evaluation of each evaluation environment, test subjects interact with the biometric system many times as it was required and both, the biometric system recognition outcomes and environmental conditions are recorded at the same time. From such results, it is possible to determine the biometric system performance (i.e. error rates and throughput rates) for the specific evaluation conditions. Furthermore, the comparison between results of the REE and the TEEs allows knowing whether the biometric system is influenced, or not, by any environmental parameter, as well as quantifying this influence. A schema of the evaluation methodology model is shown in Figure 2.

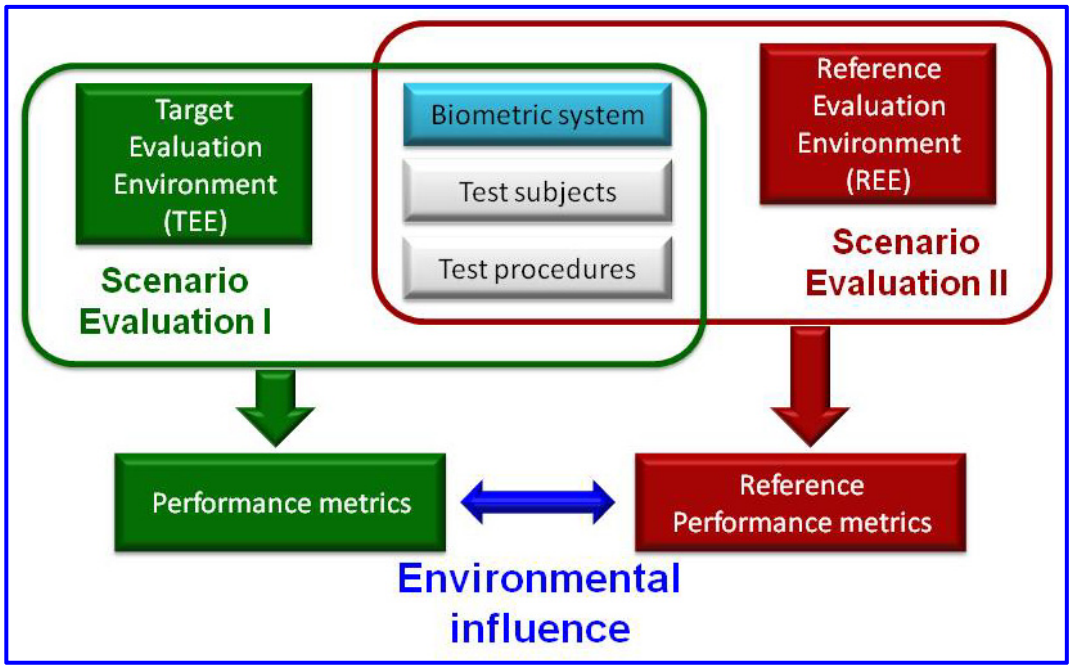

Figure 2. Evaluation model for environmental testing of biometric systems

As it has been explained previously, each evaluation environment is defined by specific environmental conditions. The evaluation methodology allows tailoring these conditions according to the objectives of the evaluation. These objectives shall consider two aspects. On one hand, the biometric system under test, its modality and the technology of its capture device to select which of the environmental parameters (i.e. which kind of environmental conditions) are of interest to the study. On the other hand, the environmental specifications

for the assessed biometric system(s), the intended operational environment, the possible extreme conditions to choose, and which values of such environmental parameters shall be assessed.

Considering this evaluation model is feasible to analyse whether a single parameter, or a combination of environmental parameters, can affect the biometric system performance. Also it is possible to deduct how the biometric system works in a particular environment, compared to the same system working in the reference environment.

For quantifying the influence of one or a combination of environmental parameters, a set of target evaluation conditions shall be determined, i.e., one evaluation condition per each value to test. In these evaluation conditions, the particular environmental parameter to assess shall be fixed to a defined value or a narrow range whereas the rest of environmental conditions (i.e. environmental parameters to control) shall be kept to a value similar to the one at the REE. Regarding this way of environmental testing, it should be emphasized the fact that each of these evaluation conditions correspond to one TEE. Therefore, the more environmental conditions to assess, the more TEEs to define, and the more evaluations to carry out. This will increase the evaluation effort considerably.

Otherwise, for analysing how the biometric system works in a specific environment, only one single target evaluation condition shall be determined. In this case, the environmental parameters shall be fixed to the corresponding value or range specified for such environment. It entails to define just one TEE for each specific environment to analyse.

5.3 Evaluation conditions specification

When a biometric system is going to be tested, the first step is to plan the evaluation. During this phase, the environmental conditions for which the biometric system is going to be evaluated shall be specified. This section addresses requirements for defining and measuring such evaluation conditions for all potential environmental parameters that can be tested during this kind of evaluations.

5.3.1 Definition of evaluation conditions

The specification of the evaluation conditions consists of determining which environmental parameters are going to be considered during the experiments. ISO/IEC 19795-1 Clause 6.4 establishes four types of controlling factors:

- factors considered part of the experiment which effects are going to be observed,
- factors considered part of the experimental conditions which are going to be controlled,
- factors out of the experiment which effects are randomized, and
- insignificant factors which are going to be ignored.

For an environmental evaluation two kinds of these factors are relevant and are considered compulsory to be defined before the evaluation:

- Environmental parameters to assess. These parameters will be part of the experiments as independent variables. It is their influence which is going to be studied. Besides, one or more fixed values or narrow ranges shall be determined for each of these parameters. These values are designated as **measuring points**.
  At least one environmental parameter to asses and one measuring point for such parameter shall be specified for the evaluation.
- Environmental parameters to control. These parameters will be part of the experimental conditions. These are environmental factors that might influence biometric performance and for this reason, it has been decided to control them. Nevertheless, they are not the target of the trial. A reference value or narrow range shall be defined for each of these parameters. This specific value (or range) is designated as **set point** and must be the same as the value defined for that parameter in the REE. It is optional to specify any environmental parameter to control, although it is recommended to specify as many as possible as to guarantee repeatability and intercomparability of the tests.

5.3.2 Type of environmental parameters

There are a lot of environmental parameters that are present when a biometric system is operating. However, not all of them affect biometric systems in the same way. As it was mentioned in section 5.2.2, it depends on two characteristics: the biometric modality and the technology of the biometric capture device.

Due to this fact, the evaluation methodology has been designed so that it is possible to select the environmental parameters to test. Nevertheless, not all environmental parameters are covered by the proposed methodology. The definition of certain factors such as vibration, mist or dust is challenging and their influence on biometrics has not been specifically mentioned in literature. Considering this fact, this work is focused on the most relevant environmental parameters addresses by ISO/IEC TR 19795-3 as influential parameters. Thus, the different types of environmental parameters that may be selected for the specification of the evaluation conditions are the following:

- Atmospheric parameters:
  - Temperature. As environmental parameter, temperature quantifies the degree or intensity of heat present in the biometric system operational environment. This parameter can affect either the system or the user biometric characteristic. It shall be defined and measured using Kelvin [K] or Celsius degrees [°C] units.
  - Humidity. This parameter quantifies the amount of water vapour in the atmosphere. It can affect either the system or the user biometric characteristic as well. The most common way to measure it is using the

relative humidity ratio. This is the ratio of the amount of water vapour in the atmosphere at a particular temperature and pressure to the maximum amount that it could hold at that temperature and pressure. Therefore, humidity shall be defined and measured using the relative humidity ratio expressed as a percentage [%]. It is important to note that there is a relationship between relative humidity and temperature. It is not possible to reach all relative humidity percentages for certain temperature values.

- Physical parameters:
  - Illumination. This parameter measures the electromagnetic radiation at different wavelengths of the electromagnetic spectrum. For users and biometric systems there are two relevant measurements: illuminance and irradiance. Illuminance quantifies the visible part of the spectrum measuring the amount of luminous flux incident on a surface. It shall be expressed in lux [lx]. Likewise irradiance quantifies the amount of radiant flux incident on a surface but covering all the electromagnetic spectrum. Moreover it shall be expressed using watts per square meter [W/m^2]. For defining and measuring this parameter, both values shall be specified in addition to their corresponding wavelength or bandwidth in nanometres [nm] in order to know the spectral power distribution.
  - Noise. This parameter quantifies the presence of loud sounds that may disturb users or make difficult to hear wanted sounds as well as to modify the sound captured in a speaker recognition system. For defining and measuring this parameter, the sound pressure level in decibels [dB] shall be specified as well as their related frequency, octave band or a one-third octave band in Hertzs [Hz] for which those levels are generated. It allows knowing the noise pressure level spectrum. Besides, it is common in noise measurement the use of frequency weighting. If any type of frequency weighting is used, this has been indicated together with the decibels (e.g. an A-weighted sound pressure level value shall be expressed as dB(A)).

Furthermore, any measurement is usually specified together with its tolerance. Therefore, environmental parameters values and measurements shall be expressed using their corresponding unit and accompanied by its tolerance.

5.3.3 Selection of the evaluation conditions

The selection of the evaluation conditions entails to determine the environmental conditions of each evaluation environment. That is to establish the environmental parameters to assess and control, as well as their related measuring and set points values respectively for the reference evaluation environment and for the target evaluation environment(s). Nevertheless, the specification of these conditions shall also consider the different phases of a biometric performance scenario evaluation, i.e. enrolment and recognition. Figure 3 shows a diagram that describes the overall process. However, according to the requirements that will be established in the following pages, most of the selected values must be similar.

Firstly, the decision of which environmental parameters must be assessed and controlled shall be done by parties involved in the evaluation. As already mentioned, this decision should be based on the biometric modality of the system under test, the target scenario where the system is to be sued, and the type of technology used by its capture device. For doing this task, it is recommended to refer to the technical report ISO/IEC TR 19795-3 which lists environmental factors that can impact biometric performance for the most relevant modalities. It is important to highlight that this decision shall be kept during the overall evaluation. In other words, the selected parameters to assess and control are the same for both the REE and the TEE(s).

Then, the particular values for all the defined environmental parameter to assess and control shall be specified. The selection of these values must conform to the requirements that are given in the text below. These requirements have been established taking into account different evaluation objectives as well as whether the intended operational environment is known or not.

A general requirement for the selection of the evaluation conditions is that if some parameters are dependent, the specification of these parameters shall be according to their dependency.

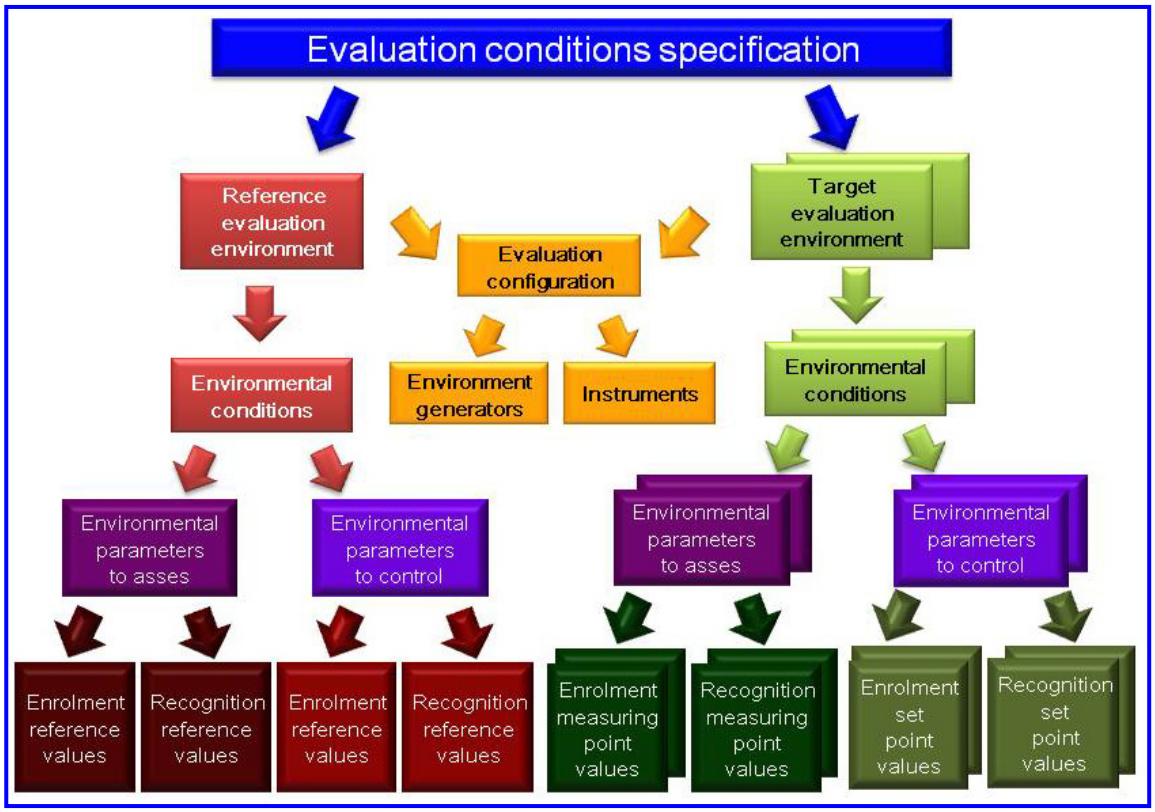

Figure 3. Evaluation conditions specification

5.3.4 Reference evaluation environment (REE)

The evaluation conditions for the REE shall be defined considering that these are the environmental parameter values or ranges under which baseline performance data will be obtained. Therefore, such evaluation conditions correspond to reference values. The test laboratory must be able to reach these reference values without any additional equipment or with equipment that do not disturb test subjects. Any factor that could affect test subjects' interactions may bias environmental testing results. In case this biasing happens, it shall be reported.

In order to establish such values, there are several possibilities: the typical values of the test laboratory, the typical values of the operational environment, a reasonable range in accordance to the biometric system specification and conventional standard conditions similar to other environmental testing methodologies. For the three first options it is not possible to determine them in a generic way as they depend on the specific laboratory and/or system under test. But the fourth one, the standard conditions, must be generically defined. Therefore, an analysis of environmental testing standards and guidelines have been performed for defining which values are going to be considered as standard conditions for each environmental parameter. The reviewed documents and the extracted information are shown in Table 6.

Table 6. Standard conditions in related standards

Environmental parameter	MIL-STD-810G (Controlled ambient)	IEC 60068-1	DEF STAN 00-35 Part 3 Issue 4	OHS Office Ergonomic Guidelines	CEN EN 12464-1	DIRECTIVE 2003/10/EC (Limits)
Temperature	23°C (± 2°C)	15°C to 35°C	15°C to 35°C (± 2°C)	21°C to 24°C (summer) 19°C to 22°C (winter)	----	----
Relative humidity	50% (± 5%)	25% to 75%	25% to 75% (±5%)	40% to 60%	----	----
Illumination	----	----	----	Common tasks: 300 lx to 400 lx Visual tasks: 600 lx	Common tasks: 500 lx to 1000 lx Visual tasks: >1000 lx Operating room: 5000 lx	----
Noise	----	----	----	55 dB(A) to 65 dB(A)	----	Exposure: 87 dB(A) Peak: 140 dB(C)

In view of these values, the standard conditions values for the different environmental parameters in biometrics have been specified as stated in Table 7.

Table 7. Standard conditions for the environmental parameters

Environmental parameter	Standard conditions value
Temperature	23 °C (± 3 °C)
Relative humidity	40% to 60% (± 5%)
Illumination	Fluorescent light - Colour temperature: 3300K to 5300K Illuminance: 300 lx to 1500 lx (± 5%) Irradiance: typical spectrum for fluorescent lamps
Noise	$L_{p,A,eq,T}$ < 65 dB(A) (±3 dB) being T= time for a user biometric transaction $L_{p,Cpeak}$ < 70 dB(C)

In case of illumination, the related standards and guidelines only provide average values for illuminance although the standard CEN EN 12464-1 [CEN'11] establishes three types of colour temperatures: cool, intermediate and warm. For this reason, the reference range for illumination has been set to average values of fluorescent light. This type of illumination has an intermediate colour temperature and it is the most common lighting mean used in offices. The spectral power distribution for the reference illumination value shall be similar to a typical spectrum for fluorescent lamps (as it is shown in Figure 4) with no peaks outside of the range between 350 nm and 850 nm.

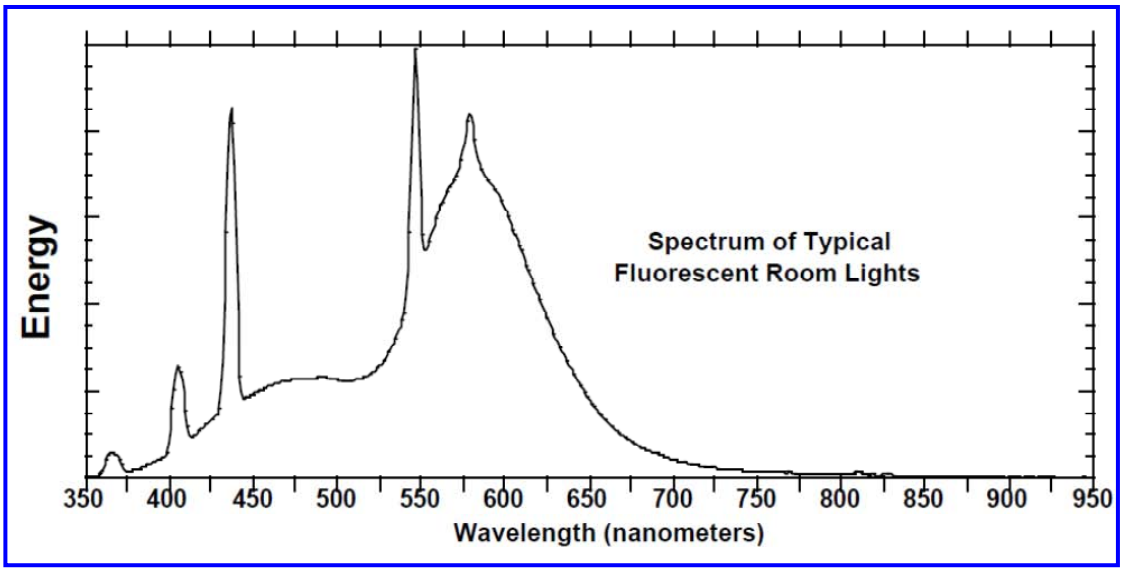

Figure 4. Spectrum of typical fluorescent lamps [ASD'99]

Likewise, for noise environmental parameter the related documents just provide the equivalent continuous A-weighted sound pressure levels for a 8 hour working day exposure. Only some regulations such as the European Union Directive 2003/10/EC [DIR'03] also define a peak value. As a result, the reference value will be defined in the same way, using the equivalent continuous A-weighted sound pressure level as expressed in equation 5. However, in this case the period of time T will be defined as the time that takes a user to complete a

biometric enrolment or recognition transaction (starting at t_1 and finishing at t_2). P_A is the A weighted sound pressure level and p_0=20 μPa.

$$L_{p,A,eqT} = 10 \log_{10} \left[\frac{\frac{1}{T} \int_{t_1}^{t_2} p_A^2(t) dt}{p_0^2} \right] \quad (5)$$

Besides, it has been necessary to define a maximum peak level in order to avoid that during that time there is any instant high pressure level that may affect the operational environment. This maximum value shall be obtained as expressed in equation 6 being p_{Cpeak} the C weighted peak sound pressure level and p_0=20 μPa.

$$L_{p,Cpeak} = 10 \log_{10} \left[\frac{p_{Cpeak}^2}{p_0} \right] \quad (6)$$

5.3.4.1 REE Enrolment evaluation conditions

The environmental conditions reference values for the enrolment depends on whether enrolment is carried out in the same operational environment that the recognition or not. Sometimes, enrolment is executed in a particular environment with the intention to obtain high quality templates. In those cases, typically the enrolment process is controlled strictly: users are under supervision and quality thresholds are severe. For those situations, it does not make sense that enrolment is covered for environmental testing and the reference values must be identical to the intended environment.

Therefore the enrolment evaluation conditions for the reference evaluation environment shall be the following:

- Standard conditions values of Table 7 when the operational environment is similar for enrolment and recognition processes, or
- Values according to the real operational environment for enrolment when the enrolment is executed in a particular controlled environment.

For those situations in which a biometric system is requested to be analyzed in a specific reference environment, which does not comply to the above mentioned standard conditions, the reference values (i.e. measuring and set points) for the enrolment evaluation conditions shall be specified previously by parties involved in the evaluation, considering the options mentioned in section 5.3.4.

Nevertheless, whatever values shall be defined it is indispensable that the test laboratory is able to reach them without any additional equipment or with equipment that do not interfere in test subjects interactions.

5.3.4.2 REE Recognition evaluation conditions

The reference values for verification evaluation conditions shall be identical to the REE evaluation conditions for enrolment except to when the enrolment is carried out in a particular controlled enrolment. In this case, the reference values shall be specified by parties involved in the evaluation according to the standard conditions of Table 7 or any of the other options given in section 5.3.4.

Again, these values shall also be defined in accordance to the requirement that the test laboratory must be able to reach them without any additional equipment or with equipment that do not interfere in test subjects interactions.

5.3.5 Target evaluation environment (TEE)

The evaluation conditions for the target evaluation environment(s) shall be defined considering that these are the environmental parameter values or ranges for which the environmental influence data will be obtained. That is, these evaluation conditions establish the measuring and set point values.

For selecting such values, two approaches may be applied. One is to base the selection on the biometric system under test and its operational range, and the other is to base the selection on the place in which the system will be located.

The first approach studies directly the biometric system performance independently where it will be located. The values are chosen from the operational range addressed by the biometric system specifications. It is suggested to analyse those values near the boundaries of the biometric system operational range in order to check whether biometric system performance is satisfactory, or not, at those questionable values.

Alternatively, the second approach checks if this biometric system is going to be affected by its actual operational environment. For this second approach values are chosen being consistent to the potential operational environment. If it is possible, it is recommended to develop a preliminary study of that environment and obtain measurements for the defined environmental factors to assess and control (e.g. the average, maximum and minimum values). If not, there are public documents and studies (e.g. NATO standard [NATO'94]) that provide tables and maps which show the environmental parameter values of different places around the world. In both situations, it is suggested to test biometric systems for the possible extreme conditions of the expected operational environment.

Furthermore, when selecting these conditions it is recommended to keep in mind that per each measuring point value, a different target evaluation environment shall be tested.

5.3.5.1 TEE Enrolment evaluation conditions

The evaluation conditions for this environment must be specified only when enrolment is covered by environmental testing, i.e. when the purpose of the evaluation includes the

comparison of the enrolment process in an environment different from the reference environment. Another possibility is to include the enrolment in the environmental testing, when the objective is to compare both enrolment and recognition processes when carried out in a reference environment against the same processes performed in a target environment. In both cases, the measuring and set point values shall be selected by parties involved in the evaluation following any of the two approaches mentioned above.

In the rest of the cases, the enrolment conducted in the TEE would be identical to the enrolment at the REE. Due to test subjects have to be enrolled once, it is probable that this process has been already done at the scenario evaluation for the REE.

5.3.5.2 TEE Recognition evaluation conditions

The evaluation conditions for this environment must be selected depending on the two possible ways this testing methodology is applied:

- For quantifying the influence of one or a combination of several environmental parameters. In this case the values have to be defined as follows:
  - Set point values shall be fixed to the standard conditions values of Table 7 or the reference values specified for the REE.
  - Measuring point values shall be selected by parties involved in the evaluation according to the two approaches explained above. It is recommended that for a predefined or observed range, at least three or four measuring points are selected: one for the minimum value, another for the maximum value and one or two more between the boundaries. For the analysis of the influence of a combination of environmental parameters, the selection of the measuring point values is similar, but considering the relationship among the combined parameters.
- For analysing how the biometric system or systems work in a specific environment. In such case the measuring and set points values or ranges shall be selected according to the values for such environment.

5.3.6 Generation and control of the environmental conditions

For performing the scenario evaluation in each evaluation environment, the evaluation conditions specified for it shall be achieved. It means that certain environmental parameters shall be modified for reaching the value or range selected for it. Then, the corresponding values or ranges shall be kept during the execution of biometric performance experiments. Both tasks shall be performed in a controlled manner which may require the use of some equipment. This equipment has been called environment generators. The requirements for these environment generators for determining that the evaluation environment has been achieved are defined below.

5.3.6.1 Environment generators to generate and control environmental conditions

The environment generators for generating and controlling the environmental parameters shall meet the following requirements:

- Environment generators shall be able to achieve the maximum and minimum value of the conditions to assess. It is recommended that they can exceed those values, in order to avoid non-lineal conditions near the generator limits.
- The resolution of the environment generators shall be appropriate in order to be able to adjust every environmental parameter values. It is recommended to have at least half of the smallest environmental parameter unit as the minimum resolution for the specified evaluation conditions.
- Environment generators shall be calibrated.
- Environment generators shall have an uncertainty lower than the one third of the tolerance specified for the environmental parameter values.
- In the case that environmental conditions are generated inside the environment generator (e.g. a climatic chamber), this environment generator shall have enough space to introduce the biometric capture device and the user's biometric characteristic.

5.3.6.2 Requirements to assure that the environmental conditions are achieved and kept

The main requirement to consider that an environmental parameter value has been reached is when its measurements are stable. That is, when this parameter is measured several times in different occasions and the result of such measurements does not change. However, the number of times that these measurements shall be done, as well as when they are going to take place, depends upon the particular environmental parameter and the environmental generator used.

Therefore, the only requirement that is established by this methodology is that the criteria used to determine the environmental parameters stabilization shall be defined and reported.

Regarding the maintenance of the evaluation conditions, is it mandatory that every environmental parameter (i.e. environmental parameters to assess and control) is kept to its fixed value or inside the range during all time that takes the scenario evaluation under such environmental conditions. Sometimes, the environmental conditions can vary sensitively due to test subject interactions. In case that any environmental parameter will be out of the specified value or range, the evaluation shall be stopped till the measuring and set point values will be achieved and are stable again. Moreover, if test subjects need to be acclimatized, the specific actions to achieve test subject acclimatization shall be carried out before continuing with the evaluation.

5.3.7 Measurement and record of the environmental conditions

Apart from generate and control environmental parameters, it is indispensable to measure and record them together with the enrolment and recognition outcomes for the purpose of environmental testing. In order to accomplish this activity, some instruments are needed. These instruments shall conform to the requirements specified in the following subsections.

5.3.7.1 Instruments to measure and record environmental conditions

Instruments used for measuring and recording the environmental parameters shall fulfil the following requirements:

- Instruments shall be able to measure a range broader than the maximum and minimum value of the corresponding environmental parameter, preferably an order with a minimum difference of an order of magnitude.
- The resolution of the instruments shall be the appropriate for recording changes. It is recommended half of the smallest environmental parameter unit specified for the evaluation conditions.
- Instruments shall be calibrated.
- Instruments shall have an uncertainty lower than the one third of the tolerance specified for the environmental parameter values.
- Instruments should have enough capacity for storing the necessary measurements or for connecting other equipment which provides such capacity.

5.3.7.2 Requirements for measuring and recording environmental conditions

The environmental parameters to be measured and recorded during the scenario evaluation in each evaluation environment are those environmental parameters chosen for being assessed. Their measurements shall be obtained and recorded at the same time that the biometric enrolment/recognition attempt is conducted and the biometric system gives the result of such attempt. These environmental parameters shall be measured in a consistent manner with the operational environment including the biometric capture device and test subjects but without affecting test subjects' interactions. The environmental parameters to control also may be measured and recorded but it is not compulsory.

For recording the environmental parameters measurements at the same time of biometric enrolment/recognition results, there are two possible methods. On one hand, the outcome of the biometric system comparison can be recorded together with the value of the environmental parameters. On the other hand, the outcome of the biometric system and environmental parameters can be recorded separately but both shall use time stamping techniques to allow the association of the values.

5.4 Fundamental requirements for planning an environmental testing of biometric systems

As it was described in section 5.2.2, environmental testing involves a biometric performance evaluation and the most proper type for the current methodology is a scenario evaluation. A scenario evaluation obtains biometric performance of a complete biometric system testing under controlled conditions which model the specific environment. Such environment is based on a real application and its target population (see section 3.5.1.2).

This section specifies all essential requirements for planning the environmental testing of biometric systems in compliance to a biometric performance scenario evaluation addressed by ISO/IEC 19795 Part 1 and 2. Basically, it establishes a generic scenario evaluation which has been adapted to analyse the influence of environmental conditions. Figure 5 shows all aspects that must be addressed and which of them have been modified for environmental testing. Although some of them do not need any modification, all of them have been described in order to provide a complete methodology.

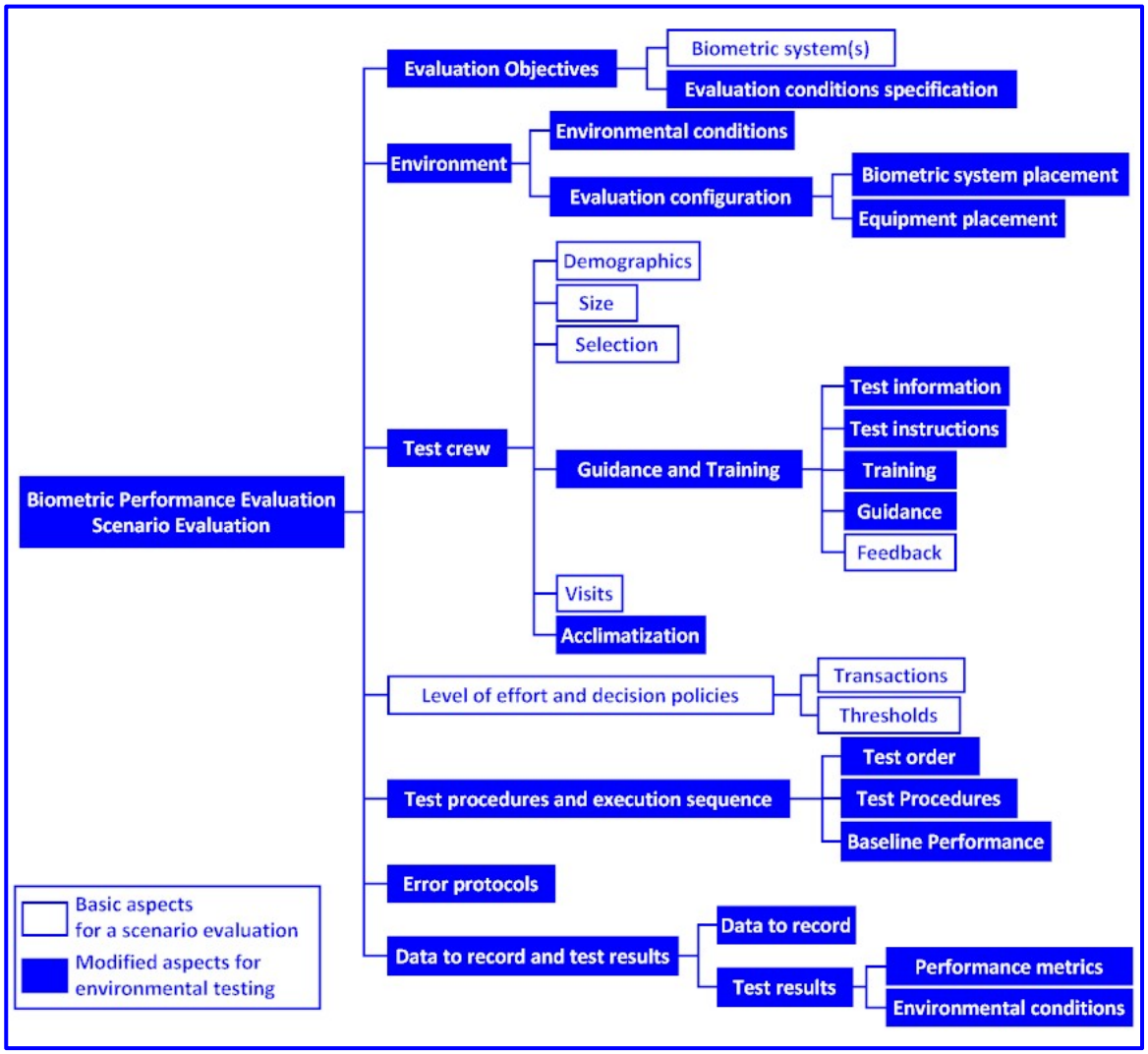

Figure 5. Scenario evaluation specification according to ISO/IEC 19795 Part 1 and 2 for environmental testing

As it will be explained below, most aspects are dependent of the intended application and the target population and shall be specified by the parties involved in the evaluation according to the evaluation objectives. In addition, other aspects shall be defined per each evaluation environment, so it is required that the test plan covers its specification for each type of evaluation environment.

5.4.1 Define evaluation objectives

For a scenario evaluation, the first step is to define the objectives of the evaluation. These shall include the following:

- A description of the biometric system(s) under test. This consists of an explanation of the biometric system(s), its modality, its capture device(s) as well as the main components that compose it. Also, it shall be described if the recognition process is based on verification (one-to-one) or identification (one-to-many) functions. For the latter, it shall be specified if it is an open-set identification or a closed-set identification too.
- A guide of the biometric system functionality. This guide must include a description of biometric functions which are implemented in the biometric system, how these functions work and their input and output parameters. This guide will be used for defining some requirements for the scenario evaluation.
- A description of the expected application including the intended operational environment (either for enrolment and recognition) as well as the target population. If it is unknown or the environmental testing is independent of the environment (i.e. it is based on the biometric system operational range specifications), it shall be clarified.
- The objective of environmental testing: to analyse the influence of one or a combination of environmental parameters or to analyse the influence of a specific environment.
- The evaluation conditions specification. A statement that claims the specific environmental parameters to assess and control and their corresponding measuring and set point values. It shall be specified in compliance to section 5.3.
- The specification of the reference and target evaluation environments to test in accordance to the evaluation conditions specification mentioned in the previous bullet. Each evaluation environment shall be described detailing the following:
    - Type of evaluation environment: reference or target.
    - Evaluation conditions for enrolment including parameters to assess and their measuring points, parameters to control and their set points and the necessary environment generators and instruments for generating, controlling, measuring and recording such environmental conditions.
    - Evaluation conditions for recognition including parameters to assess and their measuring points, parameters to control and their set points and the necessary environment generators and instruments for generating, controlling, measuring and recording such environmental conditions.

5.4.2 Operational environment

The operational environment for a scenario evaluation consists of two aspects: the environmental conditions and the evaluation configuration. Both aspects are dependent on the particular evaluation environment to be tested. Therefore, the test plan shall include a definition of them for each evaluation environment that conforms to the following requirements.

5.4.2.1 Environmental conditions

The environmental conditions for each scenario evaluation shall be fixed to the measuring and set point values defined for the specific evaluation environment as well as for the specific evaluation process (i.e. enrolment or recognition) that is going to be tested at every moment. For doing that, requirements addressed in section 5.3.6 to generate, control and maintain such evaluation conditions shall be satisfied.

5.4.2.2 Evaluation configuration

The operational environment also shall be specified in terms of where biometric system and the necessary equipment are located. For planning both issues, the following requirements shall be met.

5.4.2.2.1 Biometric system placement

If possible, the biometric system under evaluation should be located in the specified evaluation configuration in a consistent manner with the target application, biometric system supplier's recommendations and that allows test subjects to interact easily. In any case, biometric system placement shall be by agreement between parties involved in the evaluation.

5.4.2.2.2 Equipment placement

Likewise, the rest of the necessary equipment shall be located in a consistent manner with the operational environment for generating, controlling, measuring and recording the environmental parameters and biometric test subjects interactions. Environment generators shall be located in such a way that generates a uniform environment; instruments shall be located in such a way that obtains ambient measurements but their locations shall affect test subject interactions as minimum as possible.

Sometimes, it may happen that due to the values of the environmental conditions, test subject interactions have to be executed inside the environmental generator (e.g. a climatic chamber). In this situation is very probable that the evaluation configuration affects in a greater extent to the biometric system performance than the environmental conditions themselves. However, it is possible to quantify the evaluation configuration influence and isolate its effects from the environmental conditions effects. The proper method for doing it will be explained in section 5.4.5.3 when describing the test procedures for establishment the baseline performance.

5.4.3 Test crew

The set of test subjects that are going to participate in a scenario testing is called test crew. It has been demonstrated that the characteristics of the test crew influence on biometric performance [DOD'98]. Therefore, people that take part in the evaluation (i.e. the test subjects) shall fulfil the requirements specified as follows.

5.4.3.1 Test crew demographic characteristics

Test subjects shall be people which have representative characteristics of the target users. That is, test crew shall be composed by a percentage of people who gender, age, ethnic origin and occupation or technical knowledge will be similar to the percentage of end users or expected end users with the same attributes. It is important to pay attention to the physiological characteristics of the target population which are relevant for the biometric trait taken (e.g. if gait is to be used, then it is important to analyze the percentage of users in the target population that may experience mobility constraints).

5.4.3.2 Test crew size

The number of test subjects that make up the test crew shall be large enough to achieve statistically significant results. The ISO/IEC 19795-1 standard establishes the 'Rule of 3' or 'Rule of 30' to calculate the number of recognition attempts that is necessary to carry out for obtaining results at specific confidence levels. Based on this number and considering other related factors like the number of visits, the number of attempts carried out per each test subject, the availability of resources and cost and time constraints, parties involved in the evaluation shall determine the test crew size.

Due to the fact that some test subjects will probably leave the evaluation at any stage, not completing all programmed visits, it is recommended to increase test crew size in around a 10%.

For testing biometric systems based on open-set identification functions, it will be indispensable to have a group of test subjects who will not be enrolled for conducting impostor transactions. This special group shall fulfil the same requirements addressed for the common test subjects excluding those requirements related to enrolment.

5.4.3.3 Selection of test subjects

The selection of test subjects shall be random in terms of not allowing to recruit test subjects for whom the ability to recognize them is previously known. Nevertheless, the selection process shall conform to demographic requirements given in section 5.4.3.1. Moreover, test subjects must not have been involved in design, development and implementation processes of the biometric system under test and/or must not have been participated in recognition algorithm training or tuning procedures.

5.4.3.4 Guidance and training of test subjects

Another relevant factor of the test crew which influence on biometric performance is the different level of habituation of test subjects. Through suitable guidance and training procedures, this level of habituation can be balanced among test subjects and its influential effects could be reduced significantly. To that end, test subjects shall be informed, guided and trained according to the following requirements.

In case that multiple biometric systems are going to be assessed, instructions, guidance and training shall be planned considering all of them.

5.4.3.4.1 Test information

Test subjects shall be informed about the evaluation process including an overview of the evaluation, its purpose, the number of times that they must attend the testing facility, the duration of each visit and other relevant information such as legal issues related to data protection or privacy policies.

Regarding the environmental conditions, people shall be informed about the evaluation conditions in which they are going to be immersed; especially if there is any extreme condition.

It is suggested to develop forms which include the complete information about the evaluation and a declaration of acceptance to participate in it. These forms shall be signed by users before turning into test subjects.

5.4.3.4.2 Test instructions

Once people have been designated as test subjects, they shall be informed about the evaluation steps and what they have to do at each step. This explanation shall be developed according to the target application and have to include the following information:

- A description of enrolment and recognition functions, how to execute them, the number of attempts, which data must be provided by test subjects and which information are the test subjects going to receive from the biometric system.
- Instructions about how to provide the biometric characteristic to the capture device considering right and non recommended actions as well as possible information given by this device.
- Any instruction related to the possible evaluation configurations, e.g. how to act in case that there are environment generators and instruments in the operational environment or acclimatization procedures.

5.4.3.4.3 Training

Before the beginning of tests, test subjects shall perform practical enrolment and recognition attempts at different evaluation configurations. These configurations shall include equipments which are going to be used during the evaluation. During these attempts, test

operators shall supervise test subjects actions and correct any mistake. This training phase shall be adapted to the skills of each test subject and it must last till test subjects demonstrate proficiency in their interactions with the biometric system.

5.4.3.4.4 Guidance

Test subjects shall be guided during training. During enrolment and recognition it depends on the target application and the objectives of the evaluation, so it shall be decided by parties involved in the evaluation. It is recommended to guide both processes if they are controlled processes subjected to supervision or attended processes. Otherwise, enrolment and recognition should not be guided.

Nevertheless, although enrolment and recognition are decided to be non-guided processes, both shall be supervised by test operators. Such test operators shall intervene at any moment if they observe certain errors. The specific errors and the related actions to perform will be described in section 5.4.6.

In any case, guidance shall be defined during the evaluation planning in a consistent manner to test instructions including points in which guidance is required, localization of test operators to provide them, and the specific guidelines that test operator shall give to test subjects. For environmental testing, such guidelines shall be adapted to the particular evaluation conditions, evaluation configuration and acclimatization activities as necessary. Therefore, it may be needed to develop specific guidelines for each evaluation environment.

5.4.3.4.5 Feedback

The last factor regarding training and guidance is the feedback. Feedback refers to the information about the process which is provided to users by the biometric system and/or the biometric capture device by means of a display, lights or sounds.

There is not any specific requirement for environmental testing about it. Just, if the biometric system and/or its capture device provide any kind of feedback to users, it shall be given to test subjects for improving their interactions in a similar way to the final application.

5.4.3.5 Visits

Visit is a concept that refers to each time that test subjects must attend to the test laboratory for carrying out evaluation activities. Regarding this aspect, ISO/IEC 19795 Part 1 and Part 2 addresses the following:

- Multiple visits allow increasing the number of recognition transactions for only a slight rise of the evaluation cost. It is easier to get that test subjects come back to the test laboratory than to recruit new test subjects.
- Several visits allows to observe the influence of factors related to users on biometric performance such as the level of habituation (which usually improves

biometric performance) or template ageing (which typically gets worse performance).

- There shall be a time separation between enrolment and recognition attempts.

Considering these circumstances, evaluations shall have more than one visit. These visits shall take place at different times. The separation interval shall be defined in compliance to the separation time between enrolment and recognition processed at the target application.

5.4.3.6 Acclimatization

Acclimatization refers to the time that takes the human body to adapt to certain environmental conditions. This time varies depending on each person, the biometric characteristic (i.e. the modality of the system under test), the environmental parameter and its value. Therefore, according to the target application and the evaluation environment to test, acclimatization procedures should be established as necessary for different environmental parameters. Each procedure shall include the following:

- times in which this approach shall be carried out,
- minimum duration of the period for acclimatization,
- mechanisms and test subject actions to achieve acclimatization, and
- criteria to consider that test subjects are acclimatized.

It is important to consider the time that takes this process when planning the evaluation. This time may increase the duration of tests and, as a consequence, it might cause tiredness and a lack of motivation in test subjects.

5.4.4 Level of effort and decision policies

Other relevant factor of a scenario evaluation is the specification of the number of times that test subjects have to interact with the biometric system and the constraints of these interactions. This aspect is referred as level of effort and decision policies and shall meet the same requirements established for a regular scenario evaluation. Once this has been specified it will be similar for all evaluation environments.

5.4.4.1 Transactions

In order to obtain performance rates, test subjects shall be enrolled and shall execute recognition transactions. These transactions shall be as follows.

- Enrolment transactions are for generating biometric references of the test subjects. So, all test subjects shall execute this type of transaction once at each enrolment evaluation conditions except for biometric systems which operation mode is an open-set identification. For those systems the special group of test subjects selected for impostor transactions must not be enrolled. Depending on the expected evaluation effort and the biometric modality such enrolment may

generate various biometric references. Each of these shall be correctly identified in order to avoid errors.

- Recognition transactions are for checking biometric recognition functions. These transactions shall be verification transactions for testing biometric systems based on verification functions and identification transactions for testing those systems based on identification functions. In any case, test subjects shall carry out two different types of recognition transactions: genuine and impostor transactions.
    - Genuine transactions. For these transactions the test subject shall be previously enrolled at the system and it shall provide his own biometric characteristic. When testing biometric system based on verification functions, the test subjects shall provide their own identifier as well. It shall be right to avoid errors. In case of closed-set identification functions, either the test subject or the test operator shall confirm whether the identified user corresponds to the test subject. In both cases the complete test crew shall execute this type of transactions.

      On the other hand, when biometric systems based on open-set identification functions are tested, genuine transactions shall be only executed by common test subjects providing just their biometric characteristic. The special group designated for performing impostor transactions, as it has not been enrolled, is expected to provide a recognition error in their genuine transactions.
    - Impostor transactions. For performing these transactions test subjects shall provide their own biometric characteristic.

      When analysing biometric system based on verification functions, all test subjects shall execute impostor transactions. In addition to their biometric characteristic, either the test subjects, the test operator, or the evaluation system (e.g. chosen randomly) must provide the identifier of other enrolled test subject. Such identifier shall be selected randomly from available templates but excluding of the candidates those identifiers that belong to templates of the particular test subject who is going to execute the impostor transaction. This is a must because it is not a good practice to conduct impostor transactions in which samples of the same test subject are compared.

      When analysing biometric system based on open-set identification functions, only the special group of test subjects shall execute impostor transactions. In this case, test subjects do not have to provide any kind of identifier.

      At last, when analysing biometric system based on closed-set identification functions, this type of transactions shall not be executed.

Furthermore, it shall be specified the number of recognition transactions that each test subject must to carry out per visit. This number shall be determined together with the number of visits and the test crew size, as a result of applying the 'Rule of 3' or 'Rule of 30', as it was

explained in section 5.4.3.2. It is important to note that both rules are dependent of the expected error rates, so the number of genuine transactions may be different to the number of impostor transactions.

Moreover, a transaction may consist of one or more number of attempts and each attempt may consist of certain number of presentations. Therefore, the maximum number of presentations per attempt and attempts per transaction shall be specified. In addition, presentations, attempts and transactions may have a limited time to be executed. Therefore, the maximum time for accomplishing a presentation, attempt and/or transaction shall be defined as well. All these settings shall be consistent with the target application.

When testing several biometric systems, it shall be decided if the number of presentations/attempts/transactions will be identical across all systems or change according to the operation of each system. This decision concerns to parties involved in the evaluation who shall assess possible effects to modify the number of presentations/attempts/transactions for biometric systems under test or the difficulty to deal with different numbers during the evaluation process.

As a general requirement, all attempts (and transactions) shall be done with disengagement from the device. In other words, test subjects shall execute the action to present their biometric characteristic to the capture device and then the action to remove the biometric characteristic from it per each attempt. It is not allowed that test subjects present their biometric characteristic to the capture device once, and keep it positioned there to carry out all attempts.

5.4.4.2 Thresholds

Some biometric systems have configuration options that let customers to select quality and decision thresholds. When it happens, these parameters shall be fixed in a consistent manner with the target application. If quality thresholds are different for enrolment and recognition processes, the corresponding parameter for each process shall be identified and reported.

5.4.5 Test procedures and execution sequence

After establishing the requirements for all elements that are involved in the evaluation, i.e. environment, test crew and biometric system, specific procedures shall be planned for conducting the scenario evaluation in each evaluation environment. Such test plan shall satisfy the following requirements.

5.4.5.1 Testing order of evaluation environments

The order of testing evaluation environments shall be random with the intention that effects like habituation or test subjects tiredness affects biometric performance as less as possible.

However environmental testing entails to conduct two scenario evaluations at least: one for the REE and another for the TEE. As the number of evaluation conditions to analyse will be higher, the number of evaluation environments and the scenario evaluations to carry out will be also higher. As a result, the time and the effort needed for the evaluation will increase significantly. Considering these circumstances, a reasonable order of the evaluation environments to test may help to reduce them.

For this reason and when there are multiple evaluation environments to analyse, it is allowed a semi-randomness in the order. This fact shall be justified properly. Reasons for a semi-random order could be:

- to minimize the time to achieve the evaluation conditions,
- to minimize the time to change the evaluation configuration,
- to minimize the period of acclimatization of test subjects (see section 5.4.3.6), or
- the availability of equipments.

When environmental testing entails the evaluation of several biometric systems, the order of executing test subjects interactions in each system under the same evaluation environment shall be random too.

5.4.5.2 Test procedures and its execution sequence in terms of visits

In addition to establish a test order for the evaluation environments, it is necessary to plan the overall evaluation. Specifically, the plan shall include visits and which tasks to be executed in each visit by test subjects.

According to requirements already stated, at the first visit test subjects shall perform training and enrolment in all the evaluation environments. Only for biometric systems based on verification functions it would be possible to carry out the first session of genuine recognition transactions at that visit. At the subsequent visits, test subjects shall just perform the different sessions of recognition transactions in all the evaluation environments. It is suggested to develop flowcharts which include the people and the roles taking part in each test activity (i.e. test operators, test subjects, etc).

Within the test procedures planning, it shall be also decided how to arrange test subjects visits. Test subjects may come to the test laboratory alone or in a group. For the former situation, evaluation environments are changed per each test subject whereas for the latter situation, all test subjects will carry out their recognition transactions before changing the evaluation environment. Again, this aspect shall be determined by parties involved in the evaluation in a consistent manner with the difficulty to install and change the configuration of the evaluation environments, the availability of test subjects and other factors that may modify the duration of the visits like training or acclimatization.

5.4.5.3 Establishment of baseline performance

Regarding test procedures, there is another aspect that must be considered for environmental testing. This is the establishment of a baseline performance. That is, the specific procedures for obtaining reference results at predefined reference environmental conditions. In general and according to the environmental testing evaluation model, these procedures consist of carrying out the defined scenario evaluation at the REE.

However, an evaluation environment consists of two aspects: environmental conditions and the evaluation configuration. The evaluation configuration (i.e. environment generators and instruments) may also affect either the way in which test subjects interact with the biometric capture device or the perception of the feedback provided by the biometric system, changing the behaviour of test subjects. Therefore, if there is a significant modification between the evaluation configuration at the REE and the evaluation configuration at the TEE(s), this modification may have a greater influence on biometric performance than environmental conditions. Regarding this fact, test procedures for the establishment of baseline performance are explained in the following paragraphs.

If the evaluation configuration is similar among REE and TEE(s), the baseline performance shall be obtained following the general requirement. That means carrying out the specified scenario evaluation at the REE for the reference evaluation conditions.

Alternatively, when the evaluation configuration varies for the different evaluation environments in such way that affects test subject interactions (e.g. TEE entails the usage of a climatic chamber), the baseline performance shall be obtained carrying out the specified scenario evaluations twice:

- One scenario evaluation for the calculation of biometric system performance reference results. This scenario evaluation shall be performed at REE for the reference environmental conditions and the reference evaluation configuration. Due to the fact that these conditions must be reached by the test laboratory without affecting test subject interactions, the evaluation configuration corresponds to the conventional configuration.
  In order to simplify further descriptions, results of this scenario evaluation are referred as "Basic Baseline".
- A second scenario evaluation for quantifying the influence of the evaluation configuration. This scenario evaluation shall be performed at the reference environmental conditions but in the target evaluation configuration, i.e. in a configuration identical to the TEE which involves environment generators and instruments. Likewise, for simplifying further descriptions, results of this scenario evaluation are referred as "Configuration Baseline".

Since both scenario evaluations are conducted under the same reference environmental conditions, any changes in biometric performance are due only to the change in configuration. As a consequence, from the comparison of the obtained results it is feasible to quantify the

configuration influence on biometric system performance. Nevertheless, it is important to emphasize that this procedure for the establishment of the baseline actually entails the evaluation of two REEs.

Furthermore, in order to quantify the influence of environmental conditions, results for the target environment shall be compared against results for the single scenario evaluation when the evaluation configuration does not change. Otherwise for quantifying the influence of environmental conditions, results for the target environment shall be compared against results for the second scenario evaluation. Nevertheless, it will be explained in detail in section 5.4.7.2.

5.4.6 Error protocols

During the evaluation, different errors can occur. The test plan has to specify actions that test operators shall accomplish to assure that errors do not affect evaluation results. Depending on the kind of errors, these actions shall be the followed:

- General errors: these errors happen when the biometric capture device does not work correctly. In this case, the test operator shall stop the evaluation and solve the problem. Once the biometric system works properly again, the evaluation can continue.
- Environmental anomalies: if test operators detect changes in the environmental conditions, they shall measure the environmental parameters and check if these are inside their specified range. If there are any parameters outside the range, they shall stop the evaluation and correct the potential problems. Once the evaluation conditions are stable and inside the corresponding range, the evaluation can resume.
- Enrolment and verification errors: if test operators detect that the test subject has introduced a wrong identifier or has presented a wrong biometric characteristic, they shall cancel the attempt/transaction, inform the test subject about the error and the particular attempt/transaction shall be repeated by the test subject.

5.4.7 Data to record and test results

The last aspect that shall be planned for the environmental testing evaluation is the information to be recorded during experiments and how to calculate test results. If the necessary data to quantify biometric performance or environmental conditions measurements are not saved, it will be not possible to obtain evaluation results. As a consequence, the effort dedicated to the evaluation will be in vain.

5.4.7.1 Requirements for recording data

Fundamental data that shall be recorded for each evaluation environment are the following:

- environmental conditions measurements,
- the outcome of the biometric enrolment or recognition attempt/transaction,
- all kind of errors, and
- any essential information for obtaining the mandatory results addressed in the next section.

The two first parameters shall be measured and recorded according to requirements addressed in section 5.3.7.2.

It is suggested to save as much information as possible related to the outcome of the biometric enrolment and recognition attempts/transaction. The more information collected, the broader the analysis of the evaluation results become. Next, recommended data to save are specified. It is important to note that it will be not always possible to record the complete list of the below mentioned data.

- For an enrolment attempt/transaction:
    - Test subject demographics characteristics who executed the attempt/transaction.
    - Biometric characteristic(s) which are enrolled.
    - Identifier assigned to the test subject.
    - Result of the enrolment process (Successful/Failed).
    - Number of presentation/attempts needed.
    - If enrolment fails, the possible cause.
    - Quality score of the biometric sample.
    - Date and time when the attempt/transaction is executed.
    - Duration time of attempt/transaction.
    - Other relevant data (e.g. settings for the enrolment such as quality and decision thresholds).
- For a recognition attempt/transaction:
    - Test subject identifier.
    - Type of attempt/transaction: genuine or impostor.
    - Biometric characteristic which is used.
    - For impostor attempt/transaction, the identifier of the test subject who presents his biometric characteristic.
    - Similarity score or successful /failed recognition or candidate list.
    - Number of attempts needed.
    - If biometric capture or acquisition process fails, the possible cause.
    - Quality score of the biometric sample.
    - Date and time when the attempt/transaction is executed.
    - Duration time of attempt/transaction.
    - Other relevant data (e.g. settings for the recognition process, such as quality and decision thresholds and/or the number of identifiers to include at the candidate list).

If in addition to these data biometric samples are saved, it will be also possible to do *offline* testing although this kind of testing is not able to reflect all the environmental conditions influential effects as it has been explained in section 5.2.

It is also recommended to record time synchronised video recording(s) of test subjects interactions for further analysis of any errors or test subject behaviour. That further analysis will be carried out *offline*, and with special emphasis when errors have occurred.

Moreover, due to the significant amount of data generated during tests, it is recommended to automate the process systems as much as possible. With automated tools and processes systems test operator‘s work becomes easier and it prevents from human errors. Therefore, evaluation ends up being more independent and reports will be generated more easily. These systems may have multiple configurations: for biometric related data it may be a part of biometric system or application, for environmental parameters it may belong to environmental generator of instruments or, in general, it may be an independent application and/or a mixed design. The test laboratory should decide the best way to save all requested data, keeping the reliability of the whole evaluation.

5.4.7.2 Test results

Once tests have been finished, biometric performance results shall be calculated for each evaluation environment and per each biometric system under test. Specifically, these results shall consist, at least, of the following measurements:

- Environmental measurements. For each environmental parameter, it shall be obtained the following values: the minimum, the maximum and the arithmetic mean.
- Performance metrics including error rates and throughput rates:
    - Acquisition and signal processing:
        - Enrolment: FTE rate, the minimum, maximum, arithmetic mean and standard deviation time that takes to carry out an enrolment attempt/transaction.
        - Recognition: FTA rate, the minimum, maximum, arithmetic mean and standard deviation time that takes to acquire the biometric sample.
    - Comparison and decision processes:
        - Only for biometric systems based on verification functions:
            - FNMR and FMR rates. These rates may be given using ROC and/or DET curves.
            - The minimum, maximum, arithmetic mean and standard deviation time that takes a comparison attempt.
    - Complete recognition process:
        - For biometric systems based on verification functions:

- FRR/FAR and GFRR/GFAR rates. These rates may be given using ROC and/or DET curves.
- The minimum, maximum, arithmetic mean and standard deviation time that takes a verification transaction.

- For biometric systems based on open-set identification functions:
  - FNIR and FPIR rates. These rates may be given using ROC and/or DET curves.
  - Identification rate. For multiple ranks, this rate may be given by means of CMC curve.
  - The minimum, maximum, arithmetic mean and standard deviation time that takes an identification transaction.
- For biometric systems based on closed-set identification functions:
  - FNIR rate.
  - Identification rate. For multiple ranks, this rate may be given by means of CMC curve.
  - The minimum, maximum, arithmetic mean and standard deviation time that takes an identification transaction.

In addition, all measurements shall be given together with the number of attempt/transactions used to obtain these measurements and their uncertainty. In case of a biometric system based on identification functions, the number of templates that takes part in the comparison process shall be provided.

Once results have been obtained for each evaluation environment, results shall be calculated for the environmental testing evaluation. Such results disclose the environmental conditions influence on biometric performance. For this purpose, each performance metric (referred as "X") shall be generated from the comparison of the TEE results against the baseline performance results (i.e. REE results).

When the evaluation configuration is similar between REE and TEE(s), global measurements shall be obtained according to the following equation:

$$X_{\text{Environmental conditions influence}} = X_{\text{Target}} - X_{\text{Baseline}} \quad (7)$$

Otherwise, when the evaluation configuration varies for the different evaluation environments and may affect test subject interactions, global measurements shall be obtained in the following way:

1. Firstly, it is necessary to isolate the configuration effects. For doing that, results of the two scenario evaluations carrying out for establishing baseline performance shall be compared as it is expressed in equation 8.

$$X_{\text{Configuration influence}} = X_{\text{Configuration Baseline}} - X_{\text{Basic Baseline}} \quad (8)$$

2. Then, the environmental conditions influence shall be calculated by means of the comparison the target evaluation environment results against the configuration baseline results.

$$X_{\text{Environmental conditions influence}} = X_{\text{Target}} - X_{\text{Configuration Baseline}} \quad (9)$$

Moreover, it is also necessary to offer additional information about the overall evaluation process such as:

- Test crew demographics composition.
- A distribution time between visits.
- Error logs and general observations about the complete evaluation process.

5.5 Fundamental requirements for executing an environmental testing of biometric systems

Once the test plan has been developed, the next step is to conduct environmental testing in compliance with such plan. A consistent set of sequential activities shall be executed by test operators and test subjects for each of the evaluation environments. These activities have been detailed in the next subsections. When the group of activities are not listed in order, it is because the order is not relevant.

5.5.1 Pre-test activities

The test laboratory shall conduct several actions prior to conduct the evaluation environmental experiments. These shall be the following:

- Examine the biometric system(s) under test and implement the essential testing support application for performing the evaluation. It shall be able to collect the specified information and shall be conformant with the levels of effort and decision policies defined.
- Develop a plan for recruiting the needed test subjects and how these people are going to be identified.
- Develop a general evaluation schedule for arranging test subjects visits.
- Implement evaluation acceptance forms, data forms and guidelines for test subjects.
- Instruct test operators about how the biometric system works, how to use the evaluation application, how to handle equipments, how to guide and train test subjects and all necessary details to carry out the evaluation
- Develop check lists and forms which allow test operators to detect and write down errors.
- Select the necessary environment generators and instruments, calibrate them if it is necessary, check their correct operation and verify the corresponding methods for saving environmental parameter measurements.

- Prepare the lay out for the biometric system and equipments. It may entail to make a particular structure to locate them.
- Prepare additional resources for the evaluation (e.g. devices for accomplishing acclimatization procedures, tools for installing the evaluation configuration, etc.)

In addition, it is recommended to perform a mock environmental testing in which one test operator has a test subject role in order to detect if something is missing or in order to check how long it takes. Sometimes, from the results obtained in this mock evaluation, it might be needed to modify the test plan.

5.5.2 Test activities

Once, everything is ready for the evaluation, test subjects interactions shall be executed in the evaluation environments. For this purpose, the actions described in the following subsections shall be carried out.

5.5.2.1 Procedures before the first visit

At the very beginning, some tasks shall be completed before the test subjects interactions. These are the following:

- Recruit test subjects giving them appointments to come to the test laboratory at least for the first visit.
- Install the evaluation configuration in which test subjects shall execute their training including biometric system(s) and equipments.
- Verify the correct operation of biometric system covering all biometric functions that is going to be tested.

5.5.2.2 First visit

During the first visit, test operators and test subjects shall execute multiple tasks in the following order:

1. Test operators shall explain test information to test subjects and test subjects shall fill in evaluation acceptance forms.
2. Test operators shall explain test subject instructions to test subjects.
3. Test subjects shall carry out practical trials at the evaluation configuration till they demonstrate proficiency in biometric system interactions.
4. When the training will be finished, test operators shall install the enrolment evaluation environment and check that all, biometric system(s), equipments and the evaluation application for recording data work satisfactory.
5. Test subjects shall execute enrolment process. If acclimatization procedures are necessary, these shall be done before test subject interactions begin. Test operators shall guide this process in accordance with the test plan. They also shall solve any error that occurs and write it down on the error logs.

6. Dismantle the evaluation environment as necessary depending on the next steps of the evaluation.
7. If test subjects shall carry out enrolment in further evaluation environments, the steps 4 to 6 shall be repeated for the rest of evaluation environments. The order shall conform to the test order established at the evaluation plan.
8. The subsequent visits shall be set if it was not done previously.
9. Test operators shall save all data collected during this visit in a safe way.

In case of testing biometric systems based on verification functions, the steps 2 to 4 described in the next section could be carried out at the first visit but only for genuine recognition transactions.

5.5.2.3 Subsequent visits

For the rest of visits, test operators and test subjects shall carry out similar tasks to the first visit excluding those tasks related to enrolment. Specifically, the order for tasks shall be the following:

1. Test operators shall remind briefly test instructions to test subjects. At least the tasks to conduct during this kind of visits.
2. Then, the first recognition evaluation environment shall be installed by test operators. They shall check that all devices (i.e. biometric system(s), equipments and the evaluation application for recording data) work properly.
3. Test operators shall assure that the specific evaluation conditions for this evaluation environment have been reached. During this time test subjects may conduct acclimatization procedures if these are necessary.
4. Test subjects shall execute the session of recognition attempts/transactions in the evaluation environment. It entails either genuine and impostor attempts/transactions. Test operators shall guide this process in compliance to the test plan. They also shall solve and write down any inconvenience that occurs. Besides, if the environmental conditions are modified due to the interaction of the test subjects, test subjects interactions shall be stopped till these conditions reach again their corresponding values. This fact may require that test subjects shall perform acclimatization procedures again.

   In case of impostor transactions for a biometric system based on verification functions, test operators shall provide the test subject with the identifier of the template which will be forged.
5. Dismantle the evaluation environment as necessary depending on the next steps of the evaluation.
6. Steps 2 to 5 shall be repeated for all the recognition evaluation environments to test following the order established at the test plan.
7. Then, test operators shall save all data generated during the visit in a safe way.

5.5.3 Post-test activities

Finally, test operators shall calculate results and develop the corresponding reports. In particular, they shall perform the following actions.

- Obtain results per each evaluation environment.
- Calculate the general results for the environmental testing evaluation comparing results from the target evaluation environments to baseline results.
- Obtain conclusions. It is recommended to analyse error logs, video recordings and any relevant information for doing this task.
- Generate the evaluation report. This report shall include all the information stated in the next section.
- Close the evaluation. It may entail tasks such as storing all relevant information according to the test laboratory policies; remove personal data in compliance to data protection laws, dismantle biometric system(s) and other equipment, etc.

5.6 Fundamental requirements for reporting an environmental testing of biometric systems

As it has been mentioned in the previous section, the last part of the evaluation is to develop a report which gathers the results and the test procedures used for obtaining them. This report shall include the information specified as follows.

- The test plan. This document shall include all aspects that have been defined in section 5.4 as mandatory aspects to be specified either for the scenario evaluation or for environmental testing.
- Any modification performed to such test plan. This modification shall be described and justified.
- Final size of the test crew and its composition.
- A description of the methods for recording biometric data related to test subjects interactions.
- Distribution time of test subject visits and how many test subjects have participated in each visit.
- For each evaluation environment:
    - The evaluation conditions (i.e. parameter to assess and control and their corresponding measuring and set point values).
    - A relation of equipments used for generating, controlling, measuring and recording environmental parameters.
    - The specific evaluation configuration by means of photographs or diagrams.
    - Test results addressed in section 5.4.7.2.
    - Errors that have occurred during the experiments in its evaluation environment.

    - Any relevant comment considering error logs for the obtained results.
- The baseline performance results shall be indicated clearly.
- General results of the environmental evaluation as well as an analysis which interprets them. It is recommended to provide graphics which include similar measurements at different evaluation conditions. These graphics are very helpful when analysing results.
- Final conclusions for the overall evaluation.

5.7 Experiments developed for validating the methodology

Once the whole methodology has been explained, this section describes different experiments that have been conducted for developing, improving and validating the proposed environmental testing methodology for biometric systems. This description has been divided in three sections. The first section describes the preliminary studies that were carried out and the first version of the methodology. Then, the second section explains the evolution of this methodology highlighting those points which were improved. Finally, the last section described the last steps and the future improvements to the proposed methodology.

5.7.1 Preliminary studies and first version of the evaluation methodology

The starting point for the development of the environmental testing methodology was the work published under the title "Changes to vascular biometric system security & performance" [SAN'09]. This work was developed to analyse which environmental conditions influence on a vascular biometric technology. For doing that, nine scenarios were tested considering three environmental conditions: temperature, humidity and illumination. The evaluation environments and the environmental condition values can be seen in Table 8 and in Figure 6. There are two illumination values for the L5 evaluation environment because this environment entailed two locations: open air and shade. These two locations have been expressed as L5 and L5X respectively.

Regarding the evaluation methodology, this study did not provide too many details because it was focused on vascular modality and factors that may affect the security level achieved by biometric systems which use this technology.

Nevertheless, based on the methodology applied in this study, a second work was done for proposing the first version of an environmental testing methodology. This was published under the title "Evaluation methodology for analyzing environment influence on biometrics" [FER'08c]. Specifically, this document formalized the testing methodology followed at the previous work but established it in a general way, i.e. considering all modalities and biometric systems based on either verification or identification functions.

Table 8. Evaluation environments tested in [SAN'09]

Evaluation environment	Description	Environmental parameters to assess(1)	Measuring points	Environmental parameters to control(1)	Set points
L1	Standard Laboratory	Temperature	28.7 °C	------	------
		Humidity	26 %		
		Illumination	3474 Counts		
L2	Fluorescent direct lighting	Illumination	2900 Counts	Temperature	28.3 °C
				Humidity	30 %
L3	Incandescent direct lighting	Illumination	3284 Counts	Temperature	32.0 °C
				Humidity	24 %
L4	Darkness	Illumination	2212 Counts	Temperature	26.8 °C
				Humidity	34 %
L5 L5X	Direct sunlight	Illumination(2)	7123 Counts 4149 Counts	Temperature	31.5 °C
				Humidity	25 %
L6	High temperatures	Temperature	61.5 °C	Humidity	6 %
				Illumination	2205 Counts
L7	Cool temperatures	Temperature	13.3 °C	Humidity	90 %
				Illumination	2821 Counts
L8	Cold temperatures	Temperature	-14.5 °C	Humidity	92 %
				Illumination	2908 Counts
L9	Extreme Humidity	Humidity	99 %	Temperature	31.3 °C
				Illumination	2887 Counts

(1) Illumination values have been measured for a wavelength of 850 nm. This is the wavelength in which a vascular biometric system works.
(2) Different illumination values for open-air/shade locations

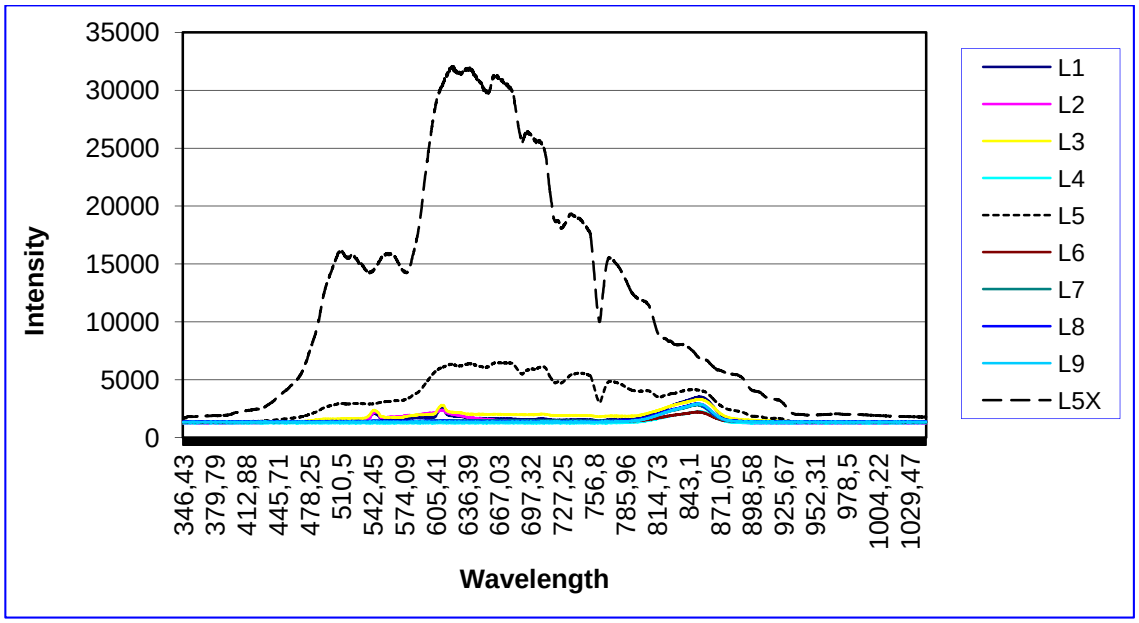

Figure 6. Spectra of the illumination for the evaluation environments tested in [SAN'09].

Specifically, this preliminary version addressed requirements regarding the following aspects:

- Environmental factors
- Tools
- Users
- Evaluation requirements
- Evaluation procedures

Although the methodology had to be improved, it was possible to analyse a vascular biometric system in all of the aforementioned evaluation environments and quantify the influence of some environmental conditions on biometric system performance. A summary of the obtained performance results are shown in Table 9. It is important to note that the vascular biometric system only provided an accept/reject decision and it was not possible to examine performance rates for different thresholds. Also, the system has fixed values for its quality and decision thresholds.

Table 9. Performance metrics results obtained in [SAN'09]

Evaluation environment	Description	Enrolment	Recognition		
		FTE	FTA	FNMR	FMR
L1	Standard Laboratory	0.0 %	26.8 %	19.6 %	0.0 %
L2	Fluorescent direct lighting	0.0 %	14.3 %	12.6 %	0.0 %
L3	Incandescent direct lighting	0.0 %	51.6 %	24.1 %	0.0 %
L4	Darkness	0.0 %	18.9 %	19.6 %	0.0 %
L5X	Direct sunlight	0.0 %	62.0 %	63.0 %	0.0 %
L6	High temperatures	0.0 %	11.8 %	22.2 %	0.0 %
L7	Cool temperatures	0.0 %	25.0%	20.4 %	0.0 %
L8	Cold temperatures	4.8 %	11.8%	22.2 %	0.0 %
L9	Extreme Humidity	0.0 %	11.8%	14.8 %	0.0 %

Analyzing this table, it can be seen that the FMR rate for the vascular biometric system is not affected by any environmental condition. However, both the FTA and FNMR rates increase considerably when the vascular biometric system has to work under illumination conditions that entail high levels of infrared light. The worst values for such rates have been obtained in L5 (direct sunlight) and L3 (incandescent direct lighting) evaluation environments. In fact, as it is explained in [SAN'09], the biometric capture device was unable to work at direct sunlight. The L5 evaluation environment was changed to a shaded location from the sun (which represents the illumination conditions shown as L5X in Figure 6, but it will be called L5 for the whole extent of this experiment, e.g. in Table 9).

Nevertheless, the most significant result was that applying the proposed methodology is was feasible to analyse and quantify the influence of environmental conditions on biometric systems performance.

5.7.2 Development of the evaluation methodology and further experiments for improving it

After the first version of the methodology, different actions were performed either to improve it or to develop a formal testing methodology to present to be presented to the biometric community.

First, several standards were analysed. On one hand, the multipart standard ISO/IEC 19795 were used to improve the overall process for planning, executing and reporting the environmental testing methodology from a biometric point of view. On the other hand, other standards that address environmental testing but for other technologies, were used for establishing requirements about the specification of environmental conditions.

Then, a refined version of the methodology was presented to ISO/IEC JTC1 SC37 WG5 as a new project for the development of an international standard. This project was included in the work plan of WG5 in 2009 with the number ISO/IEC 29197. The work titled "Environmental testing methodology in biometrics" [FER'10e] describes the scope and the contents at that time of this standard.

Since that time, many comments and contributions have been provided to the ISO/IEC JTC1 SC37 WG5 through the Spanish subcommittee AEN/CTN 71 SC37. At the same time, the feedback provided by experts from different nations who take part at the ISO/IEC JTC1 SC37 WG5 meetings has been used for conducting new experiments as well as for making progress on its development.

Some of the most important contributions were two ideas that came from experts of the US National Body. The first one was that for quantifying the environmental conditions influence on biometric performance it will be essential the establishment of a baseline performance. The second one was that this baseline shall be able to measure other possible influential effects such as the evaluation configuration. Experts thought that the fact of interacting with a biometric system placed inside a climatic chamber may modify biometric performance in a greater extent. Both ideas are interesting but needed to be matured. In addition, its incorporation to the evaluation methodology entails significant modifications compared to previous versions.

Therefore, an experiment was conducted to evolve them and to add to the corresponding requirements to the standard. In particular, this experiment consisted of testing several biometric systems in a specific environment in comparison to a reference environment with the intention of:

- defining requirements and procedures for the establishment of a baseline performance, and
- analysing whether different evaluation configurations affect biometric performance or not. It also involved the definition of methods for measuring such

influential effects as well as for isolating them to the environmental conditions influence.

Furthermore, this experiment was planned including more than one biometric system with the purpose of supplementing the methodology with requirements related to the fact that various biometric systems are tested at the same time. This entails the specification of identical policies for enrolment and verification processes for all the systems, the definition of certain test order and other actions that must be covered by the testing methodology.

In the following paragraphs, a summary of the overall experiment was described together with the obtained results. It is important to emphasize that the complete plan and report documents are not available for general public but some initial results was published in the work titled "Establishment of baseline performance for "end to end" biometric system evaluations" [FER'10c]. For confidentiality reasons, the specific biometric systems tested cannot be revealed. Any reference to them will be by a number. Also, these systems have been hidden in photographs after a label which indicates its number.

For this evaluation three fingerprint biometric systems based on verification functions were tested. The objective of the environmental testing was assessed biometric performance when systems are working in a typical hot humid environment, i.e. 40± 2°C of temperature and 60 ± 5% of relative humidity generated artificially in a test laboratory, in comparison to a the common environment of the laboratory, i.e. 26 ± 2°C of temperature and 40 ± 5% of relative humidity. In this case, two environmental parameters were assessed: temperature and humidity and one environmental parameter was controlled: illumination. This controlled parameter was selected because the three systems have a biometric capture device which uses optical technology. This type of sensor might be influenced by illumination. Its value was fixed to the fluorescent light which has the laboratory in addition to a cold light lamp for a better illumination. Besides, enrolment was considered a controlled process carried out in an environment identical to the predefined reference environment, i.e. the test laboratory.

Regarding these objectives and in compliance with requirements to specify the evaluation conditions, there were two evaluation environments:

- REE (called for the experiment "Laboratory")
    - Environmental conditions:
        - Enrolment: Values according to the real operational environment (See section 5.3.4.1).
        - Verification: Values according to the predefined reference evaluation environment (See section 5.3.4.2).
    - Evaluation configuration: laboratory
- TEE (called for the experiment "Chamber On")
    - Environmental conditions:
        - Enrolment: Values identical to the REE due to this process was going to be a controlled process (See section 5.3.5.1).

- Verification: Measuring and set point values of the specified hot humid environment (See section 5.3.5.2).

  - Evaluation configuration: inside a climatic chamber

Nevertheless, for quantify the evaluation configuration influence was necessary to add one evaluation environment more. It was called for the experiment "Chamber Off" because the chamber was switched off at this environment.

- Additional REE (see section 5.4.5.3).
  - Environmental conditions: identical to the aforementioned reference evaluation environment
  - Evaluation configuration: inside the climatic chamber

The establishment of the baseline performance involved to analyse the two reference evaluation environments. To distinguish them the first one, i.e. Laboratory, refers to the designated as "Basic Baseline" at the above detailed methodology, and the second one, i.e. Chamber off, refers to the designated as "Configuration Baseline".

Considering the aforementioned requirements, the particular environmental conditions to test and its corresponding evaluation environment were established as it has been summarized in Table 10 for enrolment and in Table 11 for verification. Moreover, those environments that were essential for the establishment of baseline performance have been indicated.

Table 10. Evaluation conditions specification for enrolment

Evaluation environment	Type of evaluation environment	Environmental parameters to assess	Measuring points	Environmental parameters to control	Set points	Evaluation configuration
Laboratory	Reference Evaluation environment	Temperature	26 ± 2 °C	Illumination	Fluorescent + Cold light(1)	Laboratory
		Humidity	40 ± 5 %			
(1) The illumination spectrum is the spectrum of Figure 7						

Table 11. Evaluation conditions specification for verification

Evaluation environment	Type of evaluation environment	Environmental parameters to assess	Measuring points	Environmental parameters to control	Set points	Evaluation configuration
Laboratory	Reference Evaluation environment "Basic Baseline"	Temperature	26 ± 2 °C	Illumination	Fluorescent + Cold light(1)	Laboratory
		Humidity	40 ± 5 %			
Chamber off	Reference Evaluation environment "Configuration Baseline"	Temperature	26 ± 2 °C	Illumination	Fluorescent + Cold light(1)	Climatic Chamber
		Humidity	40 ± 5 %			
Chamber on	Target Evaluation environment	Temperature	40 ± 2 °C	Illumination	Fluorescent + Cold light(1)	Climatic Chamber
		Humidity	60 ± 5%			
(1) The illumination spectrum is the spectrum of Figure 7						

The next figure shows the illumination spectrum for all evaluation environments either for enrolment or for verification. This spectrum comes from the combination of the fluorescent illumination of the laboratory as well as an additional cold light lamp used for improving the visibility to test subjects.

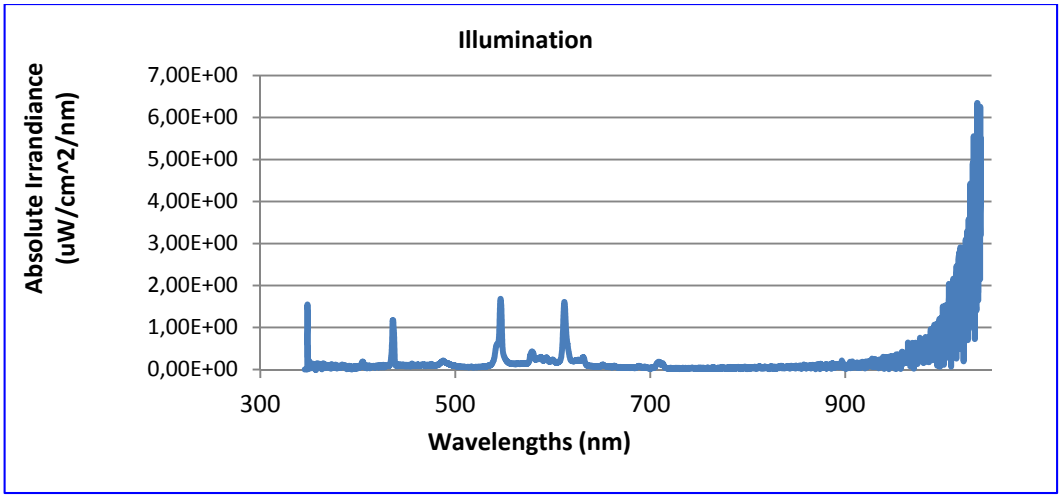

Figure 7. Illumination for all evaluation environments

Regarding the evaluation configuration, it was necessary to use two kinds of environmental generators and two instruments. The specific equipments were the following:

- Environmental generators:
  - Temperature and humidity: a climatic chamber which is able to generate a range of temperature from -70°C to 100°C and a range of relative humidity from 10% to 95% of relative humidity. Its resolution is 0.1°C for temperature and 0.1% for relative humidity whereas its accuracy is ±0.5°C for temperature and ±2% for relative humidity. In order to hold the environmental parameters to the fixed value, this chamber has been provided with a glass with a hole. Test subjects will have to interact with the biometric system through such hole.
  - Illumination: cold light lamp together with the fluorescent illumination of the laboratory.
- Instruments:
  - Temperature and humidity: a thermo hygrometer which is able to measure at the same time temperature and relative humidity. Its measurement range is from -20°C to 60°C for temperature and from 10% to 95% for relative humidity. Its resolution is 0.1°C for temperature and 0.1% for relative humidity whereas its accuracy is ±0.5°C for temperature and ±3%for relative humidity. This instrument has a sampling rate of two samples per second.
  - Illumination: spectrometer with an integrating sphere which is able to measure light intensity between 200nm and 1100nm. It has a sensitivity of

up to 130 photons/count at 400nm and 60 photons/count at 600nm. It resolution is 0.3 FWHM and its integration time is from 10µs a 65s.

Figure 8 shows the distribution of these equipments at the two kinds of evaluation environments including the situation of biometric systems. It is important to highlight that the distance between the lamp and the biometric systems was similar in both evaluation environment as well as the height of the location for the biometric systems.

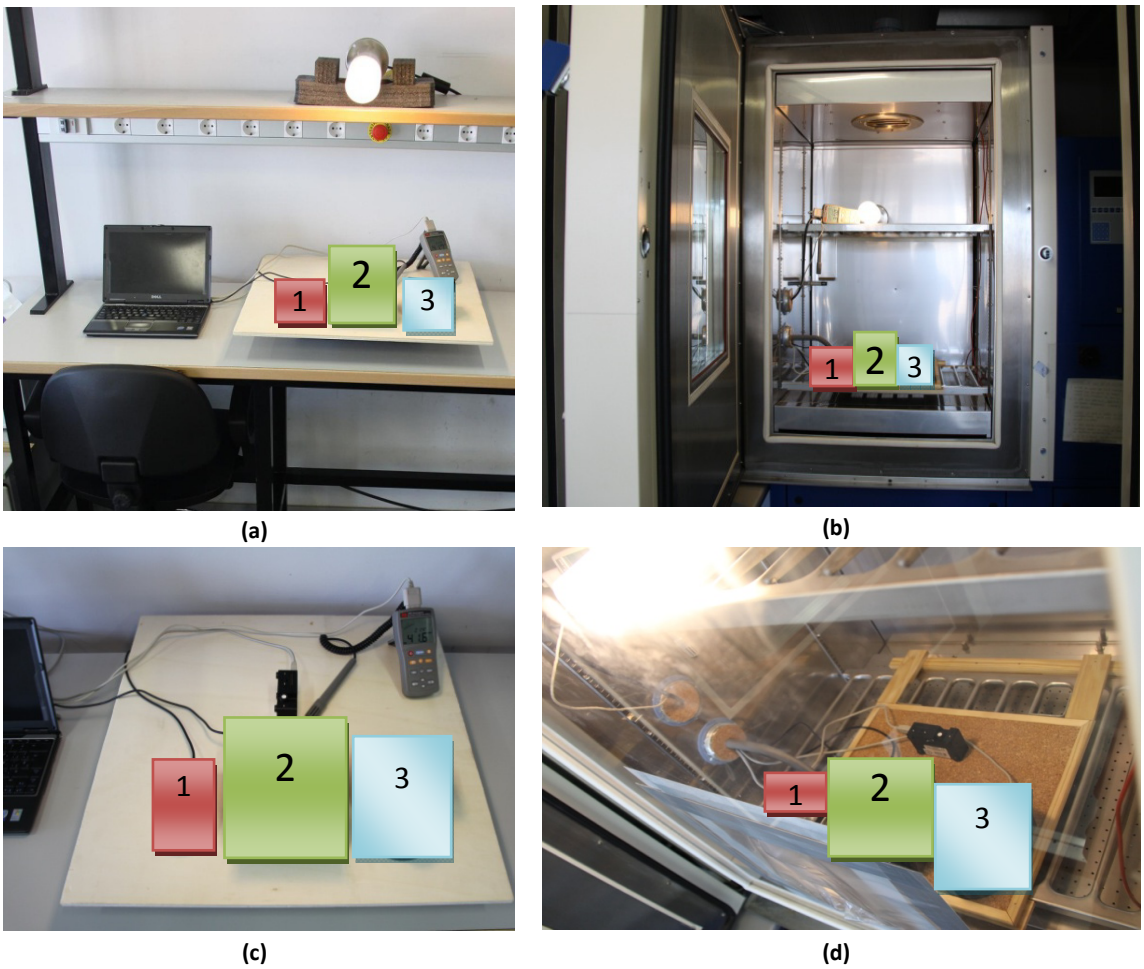

Figure 8. Evaluation configuration for the reference evaluation environment (a) front view and (c) top view as well as for the target evaluation environment (b) front view and (d) top view

Considering the biometric performance scenario evaluation, a summary of the most important characteristics is described as follows.

- Test crew: analysing the error rates claimed for each of the three biometric systems, the minimum error rate for this experiment was estimated in 0.1%. There is one system that achieves lower rates but it entails a considerable effort for the purpose of the evaluation. Therefore, applying the Rule of 3 for a confidence level of 90%, 2,000 independent biometric comparisons shall be executed at minimum. That means that it is necessary 2,000 individuals for performing the evaluation. Likewise, the recruitment of that number of the test subjects requires an excessive effort for the intention of the evaluation so, the

quantity of comparisons will be achieved but using samples of the same person (not totally independent) and a limited number of users (9 test subjects).

Number of comparisons = 9 test subjects x 4 fingers x 2 hands x 3 attempts x 5 transactions x 2 visits = 2.160 comparisons

Exactly, this group of test subjects is made up of 7 men and 2 women between 26 to 30 years old. Furthermore, all test subjects are habituated users and most of them have already used some of the biometric systems in similar evaluations [FER'10b]. Even though, a little explanation about the evaluation was provided to the test crew. Also, test subjects were instructed for interacting with the biometric systems at the climatic chamber configurations.

- Level of effort and decision policies: Test subjects have three attempts for enrolment. If they are enrolled at the first attempt, it will not be necessary to perform the remaining attempts. The maximum time per attempt will be 10 seconds. Moreover, they have to carry out three attempts per each recognition transaction and the maximum time will be also 10 seconds.

  Moreover, the quality and decision thresholds for those transactions were the fixed level that has each biometric system.
- Test procedures: The number of visits decided was two visits performed in different weeks. At the first visit, all test subjects were enrolled at the laboratory configuration and had to execute 5 genuine and 5 impostor transactions in each evaluation environment. During the second visit, they had to execute the same number of genuine and impostor transactions also in all the evaluation environments. The evaluation environments were ordered randomly in both visits.

  One of the most important aspects about the test procedure was the method to change evaluation conditions and the evaluation configuration. This method was planned as follows:
  1. Place measuring instruments and check that they work adequately.
  2. Establish controlled illumination. The cold light lamp had to be moved from the laboratory configuration to the climatic chamber and vice versa. Then, a light measurement was taken in order to test that it was installed correctly.
  3. Place biometric systems.
  4. Check that the biometric systems and the evaluation application work.
  5. Generate temperature and relative humidity when it was needed.
  6. Check if the corresponding parameters have been reached.
  7. Perform test subjects' interactions recording the needed information.
- Errors protocols were similar to the ones that have been addressed in section 5.4.6.
- Data to be recorded and test results. The information recorded was the essential information for obtaining the compulsory biometric performance results for verification systems as stated in section 5.4.7.2.

Once the test plan was finished, biometric systems, evaluation environments and test subjects were prepared for conducting the evaluation according to the test plan. It was necessary to develop an application for recording all data generated during the evaluation. When everything was ready, test subject visits begun. During the following days, all test subjects came to the laboratory for conducting the stipulated two visits. The average separation time for test subjects was 29 days. The execution of the test subjects' interactions took a total of 36 days although the effective time was 12 days. From this experiment, the following results were obtained.

In relation with the environmental conditions, the arithmetic mean off all measurements recorded for the different attempts are shown in Table 12. Illumination has a spectrum identical to the one shown in Figure 7.

Table 12. Measurements of environmental conditions

Biometric function	Evaluation environment	Temperature (°C)	Relative Humidity (%)
Enrolment	Laboratory	25.6 ± 0.28	38.6 ± 1.49
Verification	Laboratory	26.7 ± 1.37	34.6 ± 3.59
Verification	Chamber Off	27.6 ± 1.35	43.8 ± 4.11
Verification	Chamber On	40.4 ± 0.37	60.38 ± 0.89

Regarding the biometric performance measurements, error rates were the following. The FTE rate was 0.0% for the three biometric systems. There were not errors for enrolment process in any of the systems. Likewise, the FTA rates were also 0.0% in all the evaluation environments for biometric systems 2 and 3. In case of biometric system 1, FTA rates were 0.64% for the laboratory, 0.55% for the evaluation environment Chamber Off and 0.82% for the evaluation environment Chamber On. Analysing this results, it can be said that the acquisition process for biometric systems 2 and 3 is not affected by the environmental conditions whereas, for biometric system 1, a small influence exists. However, this influence is not caused by configuration effects. In fact, a lower FTA rate was obtained for the Chamber Off evaluation environment.

Considering other error rates, just the ROC curves of FNMR and FMR rates in all evaluation environments for the three biometric systems is presented in Figure 9. The rest of error rates (i.e. FRR, FAR, GFRR and GFAR) were derived from FTE, FTA, FNMR and FMR error rates. Depending of the outcomes given by the biometric systems, the ROC curve is a curve (when the system give backs a similarity score) or a point (when the system give backs an accept/reject decision). In case of biometric system 2, it was possible to select five different decision thresholds. This is the reason why the ROC curve for such biometric system is composed of five different points.

In Figure 9, it can be seen how the influence of environmental conditions is different among biometric systems. In a hot humid environment, biometric performance for systems 1 and 2 will be reduced whereas it will be higher in case of systems 3. On the other hand, there

are a difference between the laboratory results and Chamber Off results. It demonstrates that the evaluation configuration affects biometric performance as well. In fact, it seems that the evaluation configuration affects in a greater extent than environmental conditions. Therefore, it will be a must to analyse those effects during the establishment of a baseline performance.

Regarding throughput rates, the average time measurements are shown in Table 13. There is not a considerable difference among evaluation environments. So, it can be said that neither evaluation configuration nor environmental conditions affects the duration of biometric systems functions.

Table 13. Average time that took test subjects interactions

Biometric function	Evaluation environment	Biometric system 1	Biometric system 2	Biometric system 3
Enrolment	Laboratory	1.90 s	5.44 s	1.40 s
Genuine Verifications	Laboratory	1.76 s	2.48 s	1.75 s
Genuine Verifications	Chamber Off	1.80 s	2.44 s	1.75 s
Genuine Verifications	Chamber On	1.81 s	2.33 s	1.83 s
Impostor Verifications	Laboratory	1.71 s	2.46 s	1.74 s
Impostor Verifications	Chamber Off	1.73 s	2.37 s	1.73 s
Impostor Verifications	Chamber On	1.73 s	2.30 s	1.83 s

5.7.1 Future of the environmental testing methodology for biometric systems

After this experiment, there have been several revisions of the methodology to get the current version presented in this Thesis. All proposed modifications have been submitted to ISO/IEC JTC1 SC37 WG5 for improving the ISO/IEC 29197 project. Such project, where the editor is the author of this Thesis, is currently in its last phases of development, expecting its publication as International Standard in the following 2 or 3 WG5 meetings.

Furthermore, the environmental testing methodology has been disseminated in other areas in which security and environmental conditions are related such as critical infrastructures. Specifically, the work titled "Operational and Security Evaluation of Authentication Systems in Critical Infrastructures" [FER'11] described an application of the methodology for testing the level of security achieved by biometric systems when a crisis situation occurs.

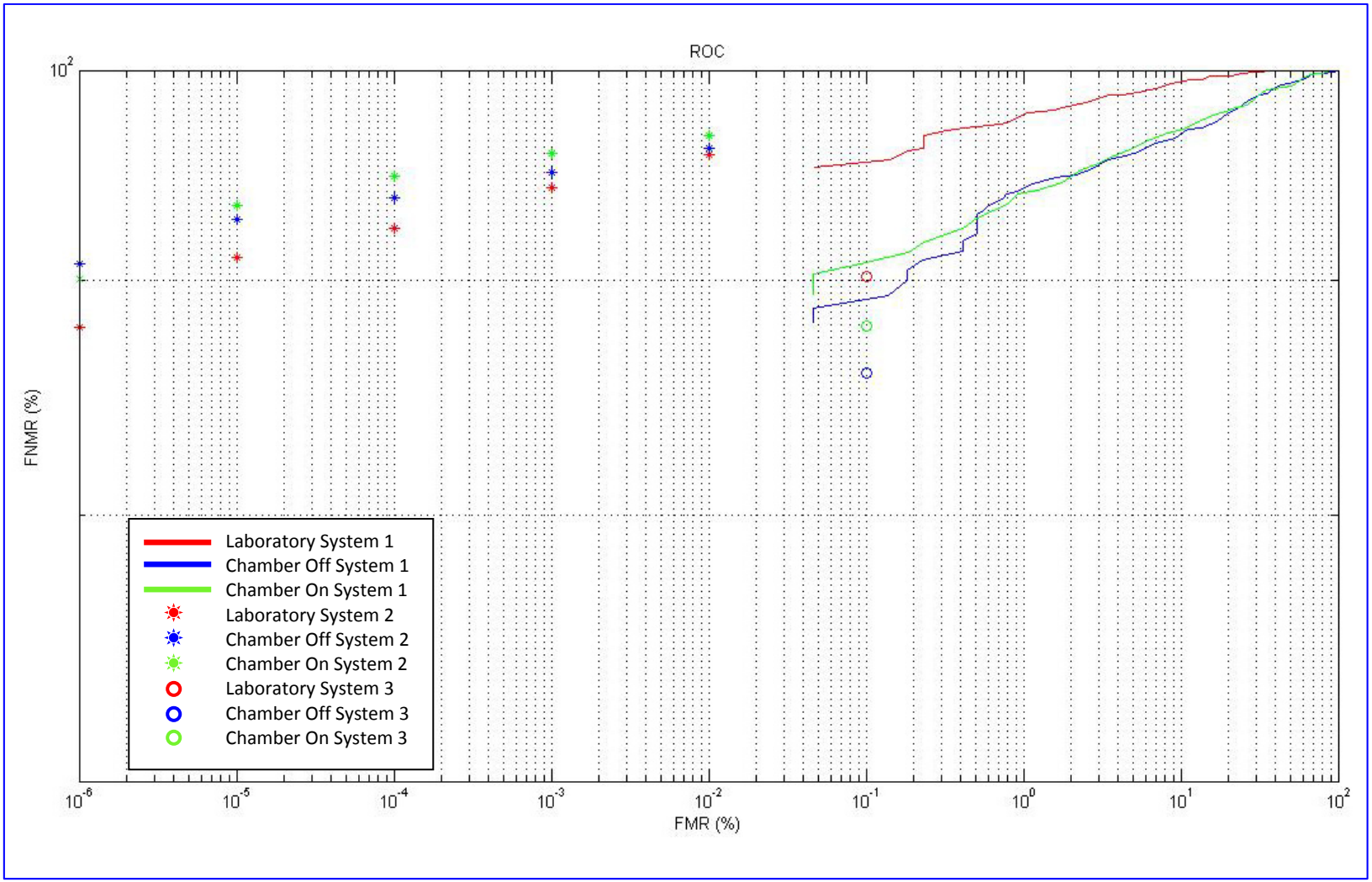

Figure 9. ROC curve for all biometric system tested

5.8 Conclusions

This chapter has presented the first main contribution of this Thesis describing an evaluation methodology to analyse the influence of ambient conditions on biometric system performance. Although this influence has been mentioned in the literature several times, no evaluation methodology had been established before the work here detailed.

Specially, this evaluation methodology has provided requirements for planning, conducting and reporting this kind of evaluation based on ISO/IEC 19795 multipart standard for planning, executing and reporting biometric performance evaluation. In particular, the following aspects have been detailed:

- Environmental conditions that may be analysed and how these conditions shall be specified for their evaluation. Also, requirements for generating, controlling, recording and reporting these conditions have been implemented. This specification has been based on environmental testing standards that currently existed for other technologies.
- Specific requirements for carrying out an environmental testing of biometric systems considering a biometric performance scenario evaluation. Exactly, additional requirements about the environment, guidance, training and acclimatization of the test crew, the sequence of execution for the different trials, error protocols, data to record and test results has been defined.
- The establishment of a baseline performance in order to accurately obtain biometric performance results for the tested environmental conditions.

Nevertheless, the proposed evaluation methodology is not only appropriate for the analysis of the influence of ambient conditions, but also for the analysis of other environmental conditions that may affect the biometric system performance such as user interaction aspects. Therefore, following a similar evaluation model, a second evaluation methodology has been proposed to analyse those effects. This evaluation methodology will be fully detailed in the next chapter.

The work in this field has provided the following set of publications:

- R. Sanchez-Reillo, B. Fernandez-Saavedra, J.Liu-Jimenez and Y-B Kwon, *Changes to vascular biometric system security & performance,* Aerospace and Electronic Systems Magazine, IEEE, 2009, 24(6), p. 4-14, 2009 [SAN'09].
- B. Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno and R. Mueller. *Evaluation methodology for analyzing environment influence in biometrics,* 10th International Conference on Control, Automation, Robotics and Vision (ICCARCV), Hanoi, 2008 [FER'08c].
- Belen Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno and O. Miguel-Hurtado. *Environmental Testing Methodology in Biometrics,* International

Biometric Performance Testing Conference (ICBP 2010), Gaithesburg, 2010 [FER'10b].

- B. Fernandez-Saavedra, F.J. Diez-Jimeno, R. Sanchez-Reillo and R. Lazarick. *Establishment of baseline performance for "end to end" biometric system evaluations,* IEEE International Carnahan Conference on Security Technology (ICCST), 2010 [FER'10c].
- B. Fernandez-Saavedra, I. Tomeo-Reyes, F.J. Diez-Jimeno and R. Sanchez-Reillo, *Operational and Security Evaluation of Authentication Systems in Critical Infrastructures,* 4th International Conference on Experiments/Process/System Modeling/Simulation/Optimization, Athens, 2011 [FER'11].
- Editor of the ISO/IEC 29197 project, which title is exactly *ISO/IEC CD 29197 Information technology -- Evaluation methodology for environmental influence in biometric system performance* [ISO'12f].

Chapter 6

Evaluation methodology for Human-Biometric system interaction testing of biometric systems

There are many other conditions that may influence the performance of biometric systems. Among them, one of the most important is the user interaction. The influence of the user interaction on biometric system performance is composed by a lot of factors. These factors may affect the acquisition process or the recognition steps.

This chapter establishes an evaluation methodology for analysing the influence of user-biometric system interaction factors on biometric systems performance [1]. Like the environmental testing methodology, this is based on the existing ISO/IEC 19795 multipart standard and considers requirements from previous studies carried out in this topic.

Initially, the chapter describes the proposed methodology including its principles, the interaction factors that should be analysed, the proper test procedures and the most relevant metrics and measurements to quantify biometric performance variations. Finally, the experiments conducted for developing, improving and validating the proposed evaluation methodology will be summarized together with the obtained results.

[1] This evaluation methodology contains similar requirements to the environmental testing methodology. Nevertheless, these will be repeated for preserving the independence of both methodologies.

6.1 Overview

The concept of human-biometric system interaction is a concept that comes from the Human-Biometric Sensor Interaction (HBSI) model. This model was recently defined by S. Elliott and E. Kukula [ELL'10] with the intention to study exhaustively all elements involved in user interactions with biometric systems and its influence on biometric system performance. Nevertheless, before describing this model and its objectives to cover with the proposed methodology, it is essential to review previous works.

Since the first biometric systems were developed, there was a common concern about the impact of user, his/her behaviour and components related to the presentation of the biometric characteristic on the biometric performance. Concepts like user acceptability, the level of habituation, whether the application is attended or not, had been traditionally claimed as factors that affect the biometric acquisition process [JAIN'98, MAN'02, WAY'04, JAI'07, ISO'07b]. However, these factors began to be studied in detail when the biometric technology became mature at two different levels.

On one hand, several modality-specific studies have been carried out analyzing different factors that affect biometric performance and/or the quality of acquired samples. Factors studied have been, for example, sensor position [NIST'06b], age [SIC'05, GUEST'06, MOD'06, FAI'11, MER'12, ERB'12], gender [MIC'08], habituation [NIST'06a, KUK'07], guidance and training [NIST'06c], instructions and feedback [COV'03, FAI'05] or the implication of having disable people in the test crew [ATHOS'05]. However, each work was conducted following its own methodology. A common methodology for analysing the influence of these factors on biometric performance did not exist.

On the other hand general, other works have been accomplished covering the following three general concepts:

- Usability. This was defined considering the definition provided by the ISO ergonomic standard 9241 [ISO'98]. That is "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use".
- User acceptance. It was defined as the demonstrated willingness within a user group to employ information technology for the tasks it is designed to support [LI'09].
- Ergonomics. It was defined according to the definition given by the International Ergonomics Association [IEA'00]. It defines ergonomics precisely as the "scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance".

The most relevant studies about usability and biometrics were developed by NIST. In 2006 this organization created a research group denominated "Biometrics and Usability" [NIST'06d]

to report their researches about usability and biometrics and to highlight to the biometric community the importance of usability in biometric applications. This group carried out some of the aforementioned modality specific studies. As a consequence, in 2008 they published a handbook [NIST'08] to provide information about how usability factors impact on biometric performance, guiding developers to design biometric products improving their usability. Regarding this handbook, it is important to say that this was written from a user-centric view. It addresses different factors to be considered when designing biometric systems like demographics characteristics (i.e. age, gender, experience and ability), guides and feedback, anthropometrics, affordance and accessibility. In addition, it proposes five usability goals to achieve (i.e. effectiveness, efficiency, satisfaction, learnability and memorability) and states different metrics to analyse how each of them affects users. However, this handbook does not establish a detailed methodology to analyse the impact of these factors on biometric performance.

In relation to user acceptance, different experiments have been conducted for analysing how users are confident using biometrics [ORC'02, ATHOS'05, HAZ'06, MOR'10]. Basically, the analysis of this aspect has consisted of performing questionnaires and surveys which ask users about aspects like privacy, safety, health risk, comfort, etc. However, these experiments have been focused on users obtaining the level of acceptability of a biometric technology or the users' attitude toward the use of biometrics. These experiments did not analyse any relationship between user acceptance and biometric performance.

Regarding ergonomics, the most relevant works have been developed by S. Elliott and E. Kukula [KUK'06, KUK'08, ELL'10, KUK'10]. In these works, they have studied the interactions of users with biometric systems in order to analyse tasks, movements and behaviours and detect potential errors. Adapting systems and processes to users reduces such errors and improve the usability of biometric systems. During these studies, they generated the Human-Biometric Sensor Interaction (HBSI) conceptual model combining the three components that are involved in human-biometric systems interactions: human beings, sensors and biometric systems as well as their resulting overlaps: ergonomics, usability and sample quality. This can be seen in Figure 10.

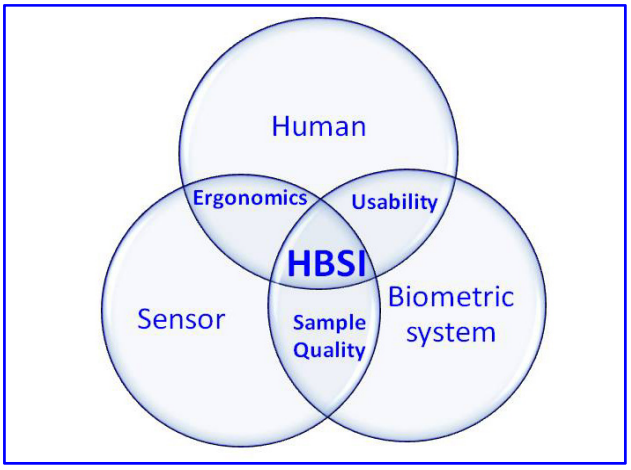

Figure 10. The HBSI conceptual model [ELL'10]

Considering this model, they propose the HBSI evaluation method which is shown in Figure 11. Basically, this method entails to calculate metrics from the different disciplines to evaluate the overall functionality and performance of a biometric system. Besides, they thought that "the traditional FTA rate (the typically usability metric) must be segmented into a more granular metrics for improving the precision of biometric performance testing". As a consequence they defined the following six new metrics [ELL'10]:

- Defective Interactions (DI): "A defective interaction (DI) occurs when a bad presentation is made to the biometric sensor and is not detected by the system".
- Concealed Interactions (CI): "CI's occur when an erroneous presentation is made to the sensor, but is not handled or classified correctly as an "error" by the biometric system".
- False Interactions (FI): "A FI occurs when a user presents his/her biometric features to the biometric system, which are detected by the system and is correctly classified by the system as erroneous due to an incorrect action, behaviour, or movement executed by the user".
- Failure to Detect (FTD): "The definition of FTD is the proportion of presentations to the sensor that are observed by test personnel but are not detected by the biometric system".
- Failure to Extract (FTX):"A failure to extract is concerned with samples from the data collection module that are unable to be processed completely". Currently, the name of this metric has been changed to "Failure to Process (FTP)" [ELL'12].
- Successful Acquisition Sample (SAS):"A successfully acquired sample occurs if a correct presentation is detected by the system and if biometric features are able to be created from the sample". In a similar way to FTX, the name of this metric has been also changed to "Successful Processed Sample (SPS)" [ELL'12].

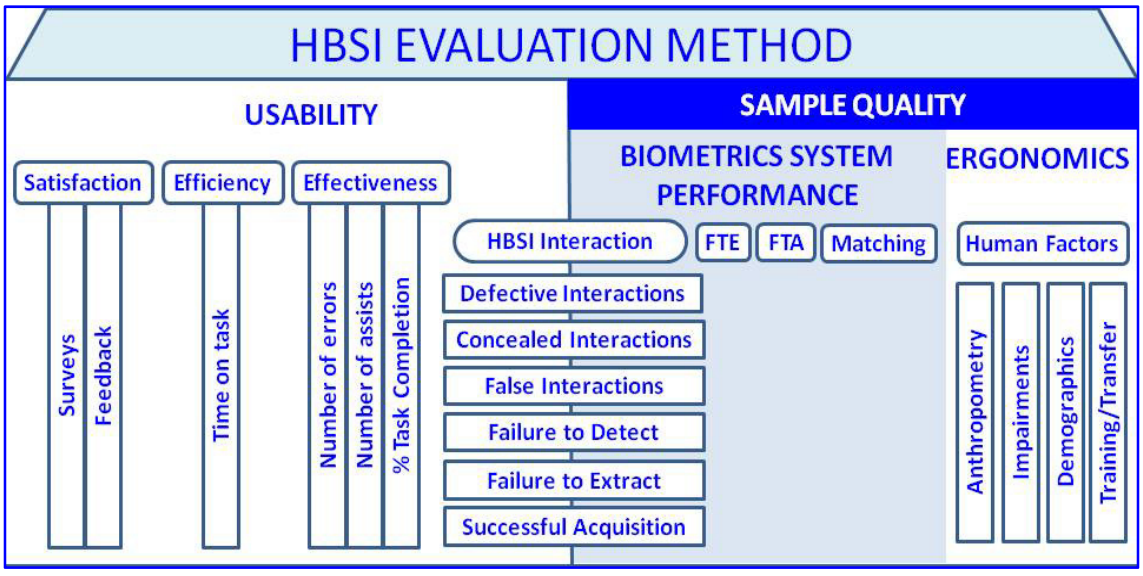

Figure 11. HBSI evaluation method [KUK'10]

However, HBSI evaluation model does not provide either which influential factors have to be analysed or the specific procedures to carry out such tests.

In view of the influential effects of ergonomics and usability factors on biometric performance and due to the lack of a formal evaluation methodology to analyse and quantify such influence, an evaluation methodology has been developed as part of the research works of this PhD Thesis. This proposed methodology is based on the HBSI evaluation method (i.e. conceptual model and metrics) and on the aforementioned NIST works. User acceptance factors have not been included because its impact on biometric performance is not direct and these factors need to be studied using a psychological perspective.

Following HBSI conceptual model and covering ergonomics and usability factors, this methodology has been named "evaluation methodology for H-B interaction testing of biometric systems". The concept "H-B interaction" refers to "Human-Biometric system interaction", where the biometric system contains both the system itself and the biometric capture device. In order to improve readability of the whole text, the shortened version of this term, i.e. H-B interaction, will be used henceforth.

This chapter describes such evaluation methodology. Specifically, the next section explains the concept of H-B interaction testing of biometric systems performance evaluations. This includes the definition of this kind of evaluations, its principles and scope. Then, next sections specify protocols and requirements that compose the methodology. Exactly, section 6.3 covers the potential factors to analyse. Section 6.4 establishes the test plan for biometric systems performance evaluation considering the previously studied factors. This test plan explanation is focused on those procedures that are different from a common biometric performance evaluation due to the analysis of the H-B interaction effects. Section 6.5 determines test execution according to the test plan and section 6.6 describes the reporting requirements. After that, the following section shows experiments accomplished to develop, test and improve the proposed methodology.

6.2 H-B interaction testing of biometric systems

H-B interaction testing is a kind of functional test in which a set of users interact with a biometric system(s) with the objective to calculate the accuracy and speed of the recognition algorithms when one or more of the following circumstances occur:

- Certain characteristics related to the biometric capture device have been modified,
- Human beings or their biometric characteristic have certain attributes, or
- Other factors related to the H-B interaction process itself have been modified.

In other words, H-B interaction testing is an "end-to-end" biometric system performance evaluation conducted considering certain usability and ergonomic factors related to the user, the biometric system or their interaction.

As it was explained in Chapter 3, there are two possible ways to carry out an "end -to-end" biometric performance evaluations: scenario and operational evaluations. However, a carefully control is fundamental for analysing ergonomic and usability factors. This fact together with the objective of this dissertation merging this methodology with CC and CEM (which claim objectivity and repeatability) makes that only scenario evaluations will be considered.

Likewise, it is indispensable that during the test the user interact with the biometric system in order to be able to observe this process. Due to these circumstances the proposed methodology only entails *online* testing. For this case, *offline* testing is not appropriate because this type of testing does not allow the analysis of users' interactions and their possible influential effects.

Furthermore and before explaining the proposed methodology and its evaluation model, it is necessary to clarify some concepts of H-B interaction that have been used for the development of the methodology. As it has been described in the previous section, there are several works developed in this area but none of them covers all essential elements that are needed for the specification of a methodology, i.e. factors to analyse and measurements to obtain. Besides of this, some of the previous concepts need to be improved. Therefore, the following sections explain in a general way the H-B interaction conceptual model used, the factors that should be analysed and the potential metrics to calculate. Then, the basic concepts for the evaluation methodology and its evaluation model will be detailed.

6.2.1 H-B interaction conceptual model

The conceptual model that has been used for this methodology has been a model based on the HBSI model developed by S. Elliott and E. Kukula previously mentioned. Nevertheless, it has been slightly modified as it can be seen in Figure 12.

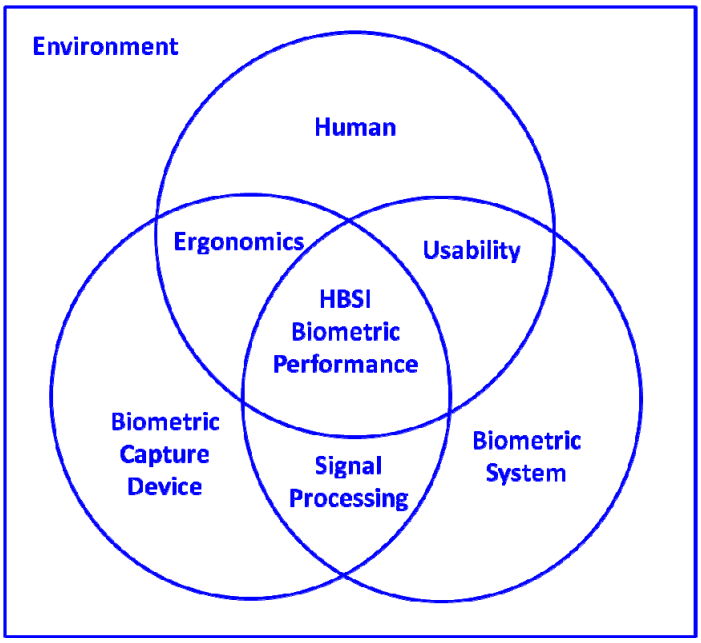

Figure 12. H-B interaction conceptual model

Firstly, the term "sensor" has been modified to "biometric capture device" according to the recent published standard ISO/IEC 2382-37 Information technology—Vocabulary—Part 37: Biometrics [ISO'12d] which established the harmonized biometric vocabulary.

Secondly, the overlap between the capture device and the biometric system has been changed to the term "signal processing" instead of "sample quality". Signal processing is a broader concept that covers all the possible processes that occur between the sensor and the biometric system such as location, segmentation, the quality improvement, feature extraction, etc.

Last but not least, it has been added the concept of "environment" as an additional element that may influence on the H-B interaction. As it was described in the previous chapter, this element may affect all HBSI components: biometric capture device, human and biometric system.

6.2.2 H-B interaction factors

Due to the fact that H-B interaction combines multiple elements (i.e. biometric system, biometric capture device, human beings, their biometric characteristics as well as the interaction between them) there are numerous factors that are subjected to be tested. Nevertheless, each of them requires the definition of specific procedures for testing them. Unfortunately it has been impossible to cover all of them within this dissertation, leaving some of them for future works.

For this reason and before specifying the methodology, next subsections present a general classification of all the possible H-B interaction factors in addition to the description of which of them factors have been covered by the proposed methodology. The factors that have been covered have been highlighted in light blue in the different tables.

6.2.2.1 Type of H-B interaction factors

This section describes a classification of H-B interaction factors. This classification includes most of the factors that have been already mentioned at NIST documents and at the ISO/IEC TR 19795 Part 3. It is also based on the components that make up the HBSI conceptual model. Considering factors that may affect each of the components, these have been classified in the three groups explained bellow.

6.2.2.1.1 Factors depending on the biometric capture device

These are factors that may influence H-B interaction because of the design, position or condition of the biometric capture device. These factors may cause that biometric sample cannot be captured or that the captured sample has a bad quality.

Particularly, the factors that compose this group are listed in Table 14. This table also includes the possible variations for some of them. In addition, an example is added to illustrate each factor and/or its possible variations.

Table 14. Factors depending on biometric capture device

Factor[(1)]	Possible variations		Example
Ergonomic design			Number of pegs in a hand geometry biometric system
Position	Height		Wall, kiosk, turnstile
	Orientation	Rotations	Place a fingerprint swipe sensor vertically or horizontally
		Inclinations	Different angles: wall or table
Condition	Damage		Scratch surface
	Dirtiness		Dust surface
(1) Factors highlighted in light blue are covered by the proposed methodology			

6.2.2.1.2 Factors depending on human beings

The second group of factors is based on users. The characteristics of individuals that are going to use the system as well as the special features of their biometric characteristics may also affect the process of capturing the biometric sample. Again, these factors may cause that biometric sample not being captured or that the captured sample does not have enough quality.

Within this group, the specific factors that can be tested are listed in Table 15. As with the previous table, this one also includes the possible variations that may have these factors, as well as an example.

6.2.2.1.1 Factors depending on human-biometric system interaction

The last group of factors are the factors which are related to interaction of the two previous components, i.e. users and the biometric system. In other words, they are factors that correspond to the interaction process itself. Likewise, these factors may affect also the process of capturing the biometric sample in a similar way that the two previous groups. These factors are shown in Table 16.

6.2.2.1 Factors which have been covered by the propose methodology

Considering the aforementioned classification, the proposed methodology only covers certain aspects because the development of a complete methodology requires a high amount of research work that has been impossible to be carried out within this single chapter. For example, environment is a factor that belongs to the third group because of this factor may affect users, biometric systems and their interactions. However and as it can be seen in Chapter 5, for analysing just the influence of this factor it has been necessary to develop a complete methodology. As a consequence, the factors which have been covered by the current methodology are the following:

- Factors that depend on the biometric capture device:
  - Position including the possible variations: height and the two types of orientation such as rotations and inclinations.
- Factors that depend on the individual:

- Temporary conditions that may affect the biometric characteristic including physical elements, behavioural aspects and chemical products.

- Factors that depend on the interaction between users and the biometric system:
  - Translations and rotations when the individual presents his biometric characteristic to the biometric capture device.

It is important to note that there is not any special reason for selecting them. The intention has been to cover a factor of each group at least.

Table 15. Factors depending on human beings

Factor[(1)]	Possible variations			Example
Biometric characteristic	Temporary conditions (It can be removed for the interaction)	Physical elements	Covered	Contact lens, glasses
			Partial covered	Hats, glasses
			Not covered but potential influence	Rings, piercings
		Behavioural aspects	Emotions	Expressions of happiness, sadness, fear
		Chemical products	Covered	Creams
			Partial covered	Make up, spots of oils, ink, paints
	Inherent conditions (It cannot be changed for the current interaction)	Short term illnesses		Loss of voice, bruises, sties, allergies, etc
		Physical appearance		Hair style, beard, moustache, losing weight
Human	Anthropometric data	Body dimensions		Tall, thin, etc
		Physical features		Eyes colour, hair colour, language accent, human laterality, etc
	Age			Children, seniors
	Gender			Men, women
	Race			Caucasian, afro-Americans, mongoloid, etc
	Experience	Habituated		User of biometrics
		Non-Habituated	With technical knowledge	Engineers, technical experts
			Without technical knowledge	Cleaning personnel
	Disabilities	Physical disabilities	Impairments	Visual, hearing, motor disable people
			Musculoskeletal disorders	Arthritis
		Mental disabilities	Cognitive	Alzheimer's disease
			Physiological	Haphephobia (Phobia of touching or being touched)

(1) Factors highlighted in light blue are covered by the proposed methodology

Table 16. Factor depending on the H-B interaction process

Factor[(1)]	Possible variations			Example
Human-biometric capture device interaction	Presentation of the biometric characteristic	Translations		Users present their biometric characteristic higher up, down, left or right than the centre
		Rotations		Roll and yaw
		Intensity		Pressure or volume
Human-biometric system interaction	Guidance	Without guidance		Non-explanation
		Non attended guidance	Visual guidance	Poster, pictograms
			Audio guidance	Sounds
			Audiovisual guidance	Video
		Attended guidance		With attendant
	Training	With training		Users receive instructions about the use of biometric system
		Without training		Users do not receive instructions about the use of biometric system
	Feedback	Without feedback		Biometric system without display, lights
		With feedback	During the process	The system indicates to the user to move forward for presenting the biometric characteristic
			At the end of the process	The system provides guidance after it is not able to capture the biometric sample
			Both	A system that includes both types of guides
Environment	Environmental conditions			Temperature, humidity, illumination, noise, etc
(1) Factors highlighted in light blue are covered by the proposed methodology				

6.2.3 H-B interaction metrics

Finally, this section describes the metrics that are going to be considered for the proposed methodology. These metrics have been based on the HBSI evaluation method developed by S. Elliott and E. Kukula. However, these have been modified as it will be shown in Figure 13.

The overall set of metrics has been organized in two main groups. The first group involves metrics which are focused on the different components which compose the model and their overlap two by two. Alternatively, the second group has been focused on metrics that provide information about the overall influence including all components on biometric system performance, i.e. the centre of the diagram.

Metrics that correspond to the first group are listed below. It is important to note that for obtaining them, it is fundamental to define each interaction in terms of tasks which are composed by actions and events. Then, the potential test subjects' actions shall be defined specifying which of them are correct, which shall be considered errors and the exact moments or events for starting and finishing counting time. In addition, particular questionnaires and surveys shall be developed and special test to check cognitive abilities. However, these definitions cannot be detailed in a generic way. For each evaluation they shall be defined considering the biometric system under test, its modality, its biometric capture device and the target application.

- Usability metrics. Usability entails the overlap between humans and the biometric system. The metrics to quantify this parameter have been divided considering the three goals which are described in the ISO definition: effectiveness, efficiency and satisfaction.
    - Effectiveness
        - Number of errors detected by test operator.
        - Number of assistance actions that test subjects need.
        - Task completion percentage.

        For each metric, the minimum, maximum, arithmetic mean and standard deviation values should be obtained.
    - Efficiency
        - Time that test subjects take to carry out an enrolment or a recognition transaction. The minimum, maximum, arithmetic mean and standard deviation values should be provided.
    - Satisfaction
        - Percentage of satisfied users. This will be measured by means of questionnaires and surveys that test subjects should fill in before, during and after they have conducted their enrolment and recognition transactions.
- Ergonomic metrics. Ergonomics involves the overlap between humans and the biometric capture device. Metrics that correspond to that group has been divided considering the two dimensions of the human beings that may affect their interactions: physical and cognitive.
    - Physical
        - Percentage of test subjects that can use the biometric capture device.
    - Cognitive
        - Percentage of test subjects that know how to use the biometric capture device.
        - Percentage of test subjects that learn how to use the biometric capture device.
        - Percentage of test subjects that remember how to use the biometric capture device.

- Signal processing metrics. Signal processing corresponds to the overlap between the biometric capture device and the biometric system. Metrics for measuring this aspect has been divided considering the two elements that are involved in the process: biometric sample and the processing capability of the different algorithms used for the recognition process.
    - Biometric sample
        - Quality metrics: modality specific metrics and quality score distribution for the obtained biometric samples.
        - Time that takes to capture the biometric sample (i.e. minimum, maximum, arithmetic mean and standard deviation).
    - Processing Capability
        - Number of segmentation errors.
        - Number of feature extraction errors.
        - Time that takes such processes (i.e. minimum, maximum, arithmetic mean and standard deviation)

Alternatively, metrics that correspond to the second group (i.e. those focused on providing information about the overall influence including all components on biometric system performance) are the traditional performance metrics (i.e. error rates and throughput rates) addressed by the ISO/IEC 19795 which were described in Chapter 3 and the HBSI error rates which come from the segmentation of the FTA rate as it was explained in section 6.1. For the calculation of the latter, it will be also necessary to specify each interaction in terms of tasks, actions and events and which correspond to each rate. Again, this definition cannot be general and must be considered the biometric system under test, its modality, its biometric capture device and the target application.

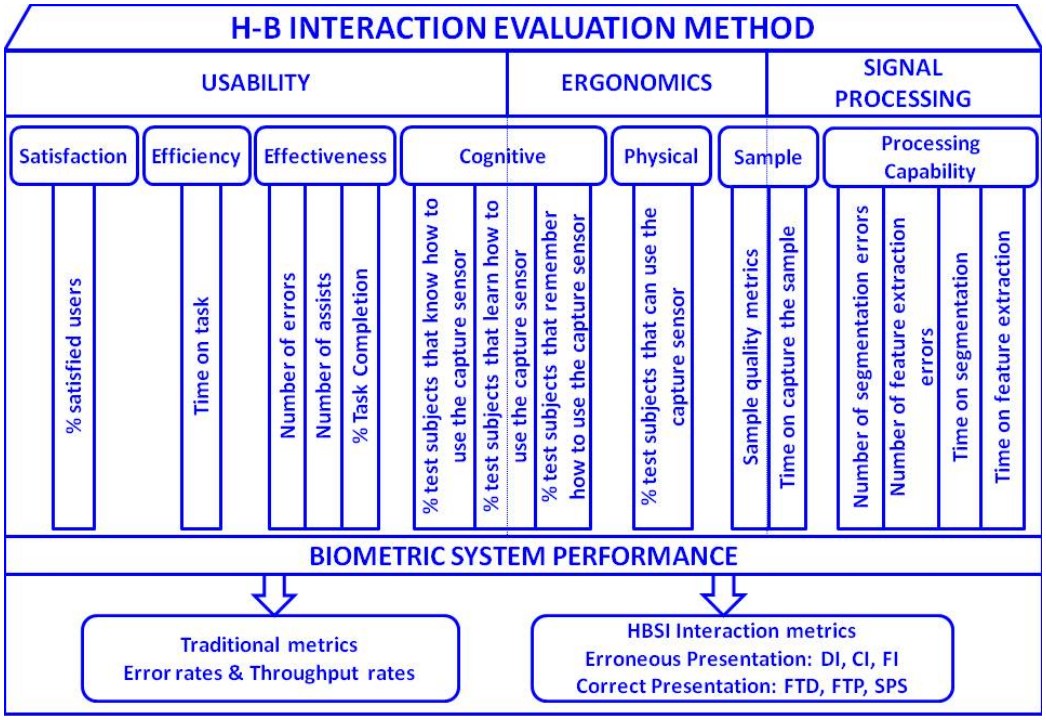

Figure 13. H-B interaction metrics

6.2.4 Basic concepts for H-B interaction testing of biometric systems

Once the scope of the H-B interaction testing methodology has been introduced, and before the description of the evaluation model that has been established, some fundamental concepts must be defined. Such definitions are provided in the next paragraphs.

H-B interaction factor

Def.: any characteristic, feature, property or condition of human beings, biometric systems or their interaction processes that may influence on biometric system performance.

These factors correspond to usability or ergonomic aspects that are inherent to the process in which users present their biometric characteristic to the biometric system.

Factor specification

Def.: detailed description of the design, feature, property or condition of a specific H-B interaction factor.

This description defines the factor and its possible variation unequivocally. Depending on the type of evaluation condition the specification can be:

- A reference specification. This is the factor specification established for reference evaluation conditions.
- A target specification. This is the factor specification defined for target evaluation conditions.

Evaluation conditions

Def.: each of the conditions which involve a different H-B interaction circumstance and which are tested for analysing their influence on biometric system performance.

There are two types of evaluation conditions[2]:

- Reference evaluation conditions (REC). These evaluation conditions entail the analysis of a reference specification for the H-B interaction factor(s) under test. For these conditions the biometric system is analysed to obtain baseline performance metrics for making comparisons.
- Target evaluation conditions (TEC). These evaluation conditions involve the analysis of the target specification for the H-B interaction factor under test. For these conditions the biometric system is analysed to obtain performance metrics for studying the influence of one or more H-B interaction factor(s), by comparing with the results obtained at the REC.

[2] REC and TEC concepts are equivalent to REE and TEE concepts used in the environmental testing methodology, respectively. However, for H-B interaction testing methodology it has been needed to define more general concepts because REE and TEE concepts refer more specifically to environmental conditions. In fact, REE and TEE concepts should be modified to REC and TEC in Chapter 5. However, it has been decided to keep them for being consistent to the ISO/IEC 29797 standardization project.

Parties involved in the evaluation

Def.: entities or organizations which are interested in the evaluation and have responsibilities in the evaluation process.

These entities are basically two: the test laboratory which is going to conduct the evaluation and the developer or customer who requests to carry out the evaluation. In case the developer is different from the customer (e.g. an end-user requesting to know the performance of a commercial product), a third entity is added to the number of parties. Test subjects are not considered a party of the evaluation although they have to take part in it.

6.2.5 Evaluation model for H-B interaction testing of biometric systems

H-B interaction testing entails to conduct two (or more) scenario evaluations: one in the reference evaluation conditions (REC) and other in the target evaluation conditions (TEC). These evaluations are identical (i.e. both shall have identical test specifications and test procedures) except for the H-B interaction factor to study. This factor will have a particular specification for each evaluation condition.

During the scenario evaluation of each evaluation condition, test subjects interact with the biometric system many times as it was required and both, the biometric system recognition outcomes as well as the test subjects' interactions are recorded. From such results, it is possible to determine the biometric system performance (i.e. error rates and throughput rates) in addition to usability/ergonomic metrics for the specific evaluation conditions. Furthermore, the comparison between results of at REC and at TEC allows knowing whether the biometric system is influenced, or not, by the analysed H-B interaction factor, as well as quantifying this influence. A schema of the evaluation methodology model is shown in Figure 14.

As it has been explained previously, each evaluation condition is specified to analyse one or a combination of H-B interaction factors. The evaluation methodology allows tailoring these conditions according to the objectives of the evaluation. These objectives may consider three general aspects:

- the design, position or condition of the biometric system and/or its biometric capture device,
- the potential users, their characteristics or the state of such characteristics, or
- parameters that may affect the interaction process such as guidance, training or feedback.

Depending on the biometric system, its capture device, the potential users and the final application, certain aspects are more critical than others. The Technical Report ISO/IEC TR 19795-3 specifies most of them for several biometric modalities. Parties involved in the evaluation shall select which of them is indispensable to analyse. It is important to emphasize that each factor to test imply the analysis of a new TEC. Therefore, the more H-B interaction

factors to asses are selected, the larger number of TECs must be tested. This may increase the evaluation effort significantly.

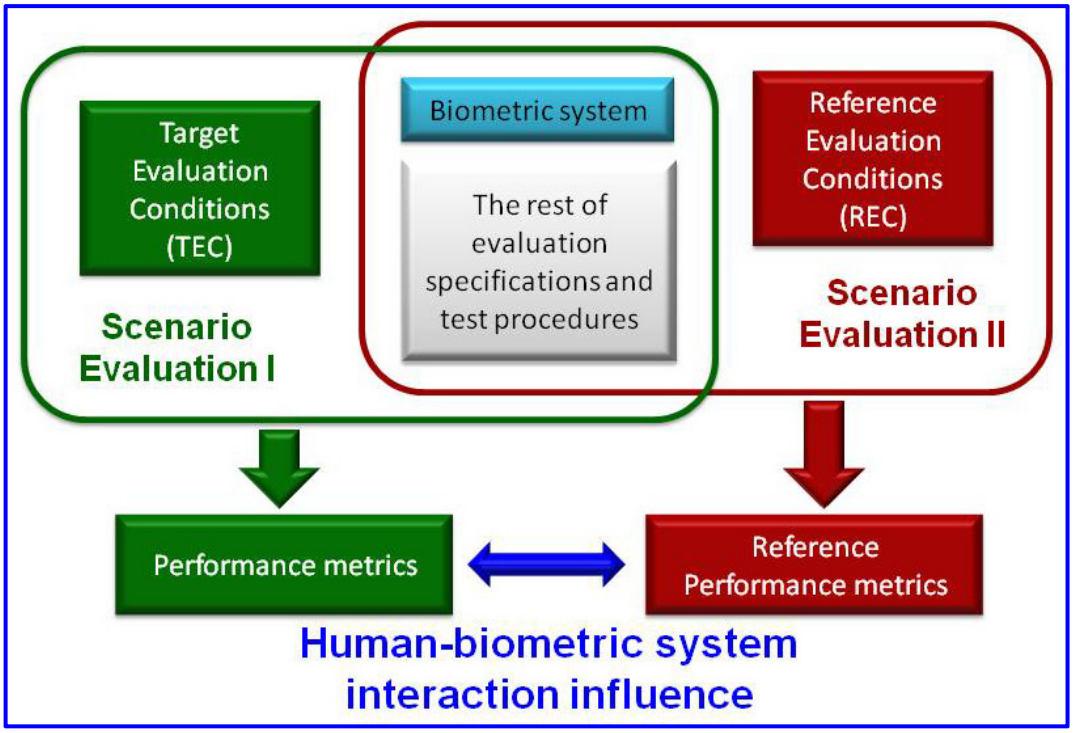

Figure 14. Evaluation model for H-B interaction testing of biometric systems

The evaluation model given is feasible to analyse whether a single factor, or a combination of H-B interaction factors, can affect biometric system performance and quantify its/their influential effects. Also, it is possible to deduct how the biometric system works considering a particular H-B interaction factor compared to the operation of the same system considering a variation of that H-B interaction factor.

6.3 Evaluation conditions specification

The first step to perform a H-B interaction evaluation is to plan the evaluation. During this phase, the H-B interaction conditions for which the biometric system is going to be evaluated shall be specified. This section addresses requirements for defining and measuring such evaluation conditions for all potential H-B interaction factors that can be tested during this kind of evaluations.

6.3.1 Definition of the evaluation conditions

The definition of the evaluation conditions consists of determining which H-B interaction factors are going to be assessed during the experiments. Considering the factors which are going to be covered for the current methodology, the different kinds of factors that may be selected for the specification of the evaluation conditions are provided in the following subsections.

6.3.1.1 Factors depending on the biometric capture device

These factors are listed in Table 17. For defining them, it is required to specify the factor to assess and the particular variation. Due to this group of factors can be quantified using unit of measure, the corresponding unit shall be provided. The precise unit for each factor has been indicated in the table. Nevertheless, additional material such as photographs, and/or diagrams may be very helpful to illustrate the definition.

Table 17. Factors depending on the biometric capture device

Factor	Possible variations		Definition
Position	Height		Distance to the ground using metric units, e.g. [m]
	Orientation	Rotations	Rotation angle expressed in degrees [°] The reference axis shall be described.
		Inclinations	Inclination angle to the horizontal in degrees [°]

6.3.1.2 Factors depending on human beings

Regarding human beings, the possible factors to study are shown in Table 18. As it can be seen in Table 15, there are multiple examples that can be considered for each factor. Therefore, an detailed description of the factor to assess and the possible variations shall be described, including multiple details about which characteristic or conditions are considered that fulfil the factor and which are not considered. Table 18 also provides the specific definition for each factor. In addition, any supplementary material such as photographs and/or diagrams may be very useful for the particular definition.

Table 18. Factors depending on human beings

Factor	Possible variations			Definition
Biometric characteristic	Temporary conditions (It can be removed for the interaction)	Physical elements	Covered	A list of possible elements and its characteristics A list of elements that are not included and its characteristics Instructions about the location of the elements
			Partial covered	
			Not covered but potential influence	
		Behavioural aspects	Emotions	A list of possible expressions and the level of expressiveness
		Chemical products	Covered	The description of the product and which body parts shall cover.
			Partial covered	

Furthermore, this group of factors may consider an additional test option, which is the requirement of having a subset of users within the test crew that do not meet the defined factor. Sometimes, it could be desired that only a percentage of the test crew satisfies the factor to assess but not the rest, or that a percentage of the test crew fulfil a factor whereas the rest fulfil a variation of the same factor. In this case, the description of the evaluation conditions shall include the percentage of the test subjects that shall fulfil the factor to assess and/or its variation and which other factors shall be fulfilled by the rest of test subjects.

6.3.1.3 Factors depending on human-system interaction

In relation to the interaction process, the possible factors to be tested are given in Table 19. To define them, the description of the factor to assess and possible variations shall be provided including the aspects indicated in the table at least. In a similar way as in the previous groups, the use of photographs and diagrams could be very helpful when defining the assessed factor.

Moreover, for this type of factors is also possible the definition that only a percentage of the interactions meets the selected factor to assess while the rest of them meet another factor or a variation of it. In this situation, the description of the evaluation conditions shall include the percentage of test subjects' interactions that shall fulfil specifically the factor to assess or a possible variation and which other factors shall fulfil the rest of them.

Table 19. Factor depending on the human-biometric system interaction process

Factor	Possible variations		Definition
Human-biometric capture device interaction	Presentation of the biometric characteristic	Translations	The reference point Direction of the allowed translation The allowed distance from the reference point expressed in metric units, e.g. [m]
		Rotations	The reference point and the reference axis Direction of the allowed rotation The rotation angle expressed in degrees [°]

6.3.2 Selection of the evaluation conditions

To select the evaluation conditions, it is needed to determine the factors to be assessed and their detailed specification for both the REC and for the TEC(s). Nevertheless, this definition conditions shall also consider the different phases of a biometric performance scenario evaluation, i.e. enrolment and recognition. Figure 15 shows a diagram that describes the overall process.

Firstly, the decision on which evaluation factors have to be assessed shall be done by the parties involved in the evaluation. As already mentioned, this decision should be based on several parameters: the biometric modality of the system under test, the type of technology used by its capture device, the target application, as well as the target population. For doing this, it is recommended to refer to the technical report ISO/IEC TR 19795-3 which lists factors that can impact biometric performance for the most relevant modalities.

Then, the particular specification for all the defined evaluation factors shall be established. The selection of this specification for each evaluation condition must conform to the requirements that are given in the text bellow. These requirements have been established considering different evaluation objectives as well as whether the intended application and the target population are known or not.

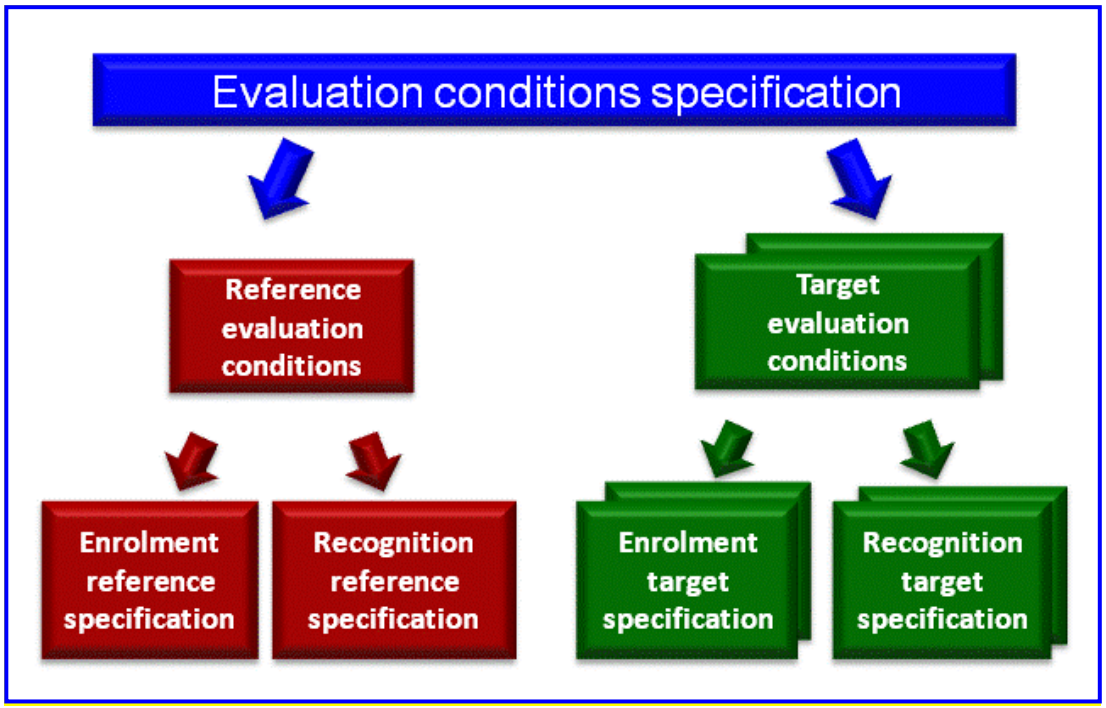

Figure 15. Evaluation conditions specification

It is important to emphasize that when selecting the evaluation conditions only the factor to analyse and the scenario evaluation aspects that involve such factor shall be specified. The rest of the evaluation conditions shall be defined according to an ISO/IEC 19795 biometric performance scenario evaluation which will be explained in section 6.4.

6.3.3 Reference evaluation conditions (REC)

The factors' specification for the REC shall be defined considering that this is the specification for which the baseline performance data will be obtained. Therefore, this shall correspond to a reference specification.

In order to establish such values, there are several possibilities based on the typical values of the target application/population or conventional conditions. For the first option it is not possible to determine them in advance, but some conventional conditions are provided for each group of factors:

- For factors depending on the biometric capture device the most proper situation is according to developers' recommendations. If these recommendations are not given, the biometric capture device shall be located considering the following:
    - Height:
        - For "desk" devices: the standard height of a table, desk or kiosk.
        - For "wall" devices: within the limits of the distance of the face to the ground, considering the average population of a defined area (e.g. the average height where the eyes of an average user are located).

  - Orientation: straight in line with the user in a standard position towards the device, without inclinations.
- For factors that depend on the human being characteristics the most appropriate reference specification is that one where test subjects do not have any component that affect or cover the biometric characteristic. That is:
  - Physical elements: test subjects must not wear any element that cover, partially cover or may affect the biometric characteristic capture process.
  - Emotions: test subjects shall not express any emotion.
  - Chemical product: test subjects must not have used any chemical product.
- For factors that depend on the interaction process the most proper reference specification is that test subjects present their biometric characteristic in compliance to developers' recommendations. If these guidelines are not provided, test subjects shall present their biometric characteristic without translations and rotations where the biometric capture device has a better response or a higher sensitivity.

6.3.3.1 REC for enrolment

The reference specification for enrolment depends on whether enrolment is carried out, either in the same conditions that the recognition process or in different ones. Sometimes, enrolment is executed in particular conditions with the intention to obtain high quality templates. In those cases, typically the enrolment process is controlled strictly: users are under supervision and quality thresholds are severe. For those situations, it does not make sense that enrolment is covered by H-B interaction testing and the reference specifications must be identical to those intended conditions (i.e. both for REC and TEC, only needing to follow the enrolment process once).

Therefore the enrolment REC shall be the following:

- Conventional conditions when the operational conditions are similar for enrolment and recognition processes, or
- Values according to the real enrolment conditions when the enrolment is executed in particular controlled conditions.

For those situations in which a biometric system is requested to be analyzed for a particular reference specification (which does not comply to the above mentioned conventional conditions), the reference specification for the enrolment evaluation conditions shall be defined previously by parties involved in the evaluation, considering the typical values for the target application and population.

6.3.3.2 REC for recognition

The reference specification for recognition evaluation conditions shall be identical to the enrolment evaluation conditions except when enrolment is carried in particular controlled

conditions. In such a case, the REC specification shall be established by parties involved in the evaluation considering the options given in section 6.3.3.

6.3.4 Target evaluation conditions (TEC)

The factor's specification for the target evaluation condition(s) shall be defined considering that these are the conditions for which the biometric performance influence will be measured. That is, these evaluation conditions correspond to the target specifications.

For selecting such values, two approaches may be applied. One is to base the selection of predefined circumstances that want to be analysed. The other is to base the selection on the target application/population and the possible circumstances that may happen.

The first approach studies directly the biometric system performance independently of the target application/population. The factors' specification is chosen according to the fixed specification which is the objective of the tests. It is suggested to analyse the most challenging circumstances in order to check whether the biometric system performance is satisfactory, or not, at questionable circumstances.

Alternatively, the second approach checks if this biometric system is going to be affected by its target application/population. For this second approach the factors' specification is chosen being consistent to the real conditions and users. If it is possible, it is recommended to develop a preliminary study of those conditions and obtain measurements for the defined location of the biometric capture device and the characteristics of the potential users). Again, it is suggested to test biometric systems for the most challenging conditions either due to the target location of the biometric system or due to the characteristics of the target population.

Furthermore, when selecting these conditions it is recommended to keep in mind that per each factor and its variation, a different target evaluation condition shall be tested. A balance between the information to be obtained and the effort (and cost) needed for the evaluation should be reached.

6.3.4.1 TEC for enrolment

The specification of the factors required for this evaluation conditions must be defined only when enrolment is covered by H-B interaction testing, i.e. when the purpose of the evaluation includes the comparison of the enrolment process for a specification of factors different from the reference specification. Another possibility is to include the enrolment in the H-B interaction testing when the objective is to compare both enrolment and recognition processes when carried out for a reference specification against the same processes performed for a target specification. In both cases, the specification for enrolment TEC shall be selected by parties involved in the evaluation following any of the two approaches mentioned above.

In the rest of the cases, the enrolment conducted in the TEC would be identical to the enrolment at the REC. Due to test subjects being enrolled once, it is probable that this process has been already done at the scenario evaluation for the REC.

6.3.4.2 TEC for recognition

The specification of factors for recognition at TEC shall be selected by parties involved in the evaluation according to the particular factors and their possible variations that are going to be tested. It is suggested to apply any of the two approaches explained above.

6.3.5 Generation of the evaluation conditions

For performing the scenario evaluation in each evaluation condition, the specification of the relevant factors shall be satisfied. There is not predefined equipment for achieving the evaluation conditions. The following tasks shall be carried out depending on the different factors:

- For factors that depend on the biometric capture device it will be essential to place the system as it has been indicated. It may require the usage of a proper structure which models the desired location.
- For factors that depend on the human beings it will be necessary to provide the test subjects with the corresponding physical element or chemical product and explain them how they shall put on or apply it respectively. In some cases it will be not possible to provide a particular element (e.g. piercings). For these cases test subjects composing the test crew shall be selected according to the defined characteristics. For emotions, it will be necessary to develop certain guidelines in order to explain to test subjects the exact expressions that they have to express.
- For factors that depend on the interaction process it will be necessary to develop guidelines for instructing test subjects about how they must present their biometric characteristics to the biometric capture device in compliance with the evaluation conditions specifications.

6.3.6 Control of the evaluation conditions

For H-B interaction testing it is required to control exhaustively that test subjects carry out their interactions according to the evaluation conditions. Therefore, test operators shall watch that the following requirements are fulfilled depending on the factors to analyse.

- For factors that depend on the biometric capture device, test operators shall check that the biometric capture device is placed as it has been specified for the evaluation conditions which have being tested at that moment.
- For factors that depend on the human beings, test operators shall check that they carry out their interactions wearing the corresponding element, doing the corresponding expression or have applied the chemical product as it was specified

respectively. It will depend on the exact evaluation conditions which have being tested.

- For factors that depend on the interaction process, test operators shall check that test subjects conduct their interactions according to the evaluation conditions specifications which have been tested.

Furthermore, for H-B interaction testing is indispensable to record test subjects' behaviour during their interactions with the biometric system. These recordings shall be done without affecting test subjects interactions but shall record the movements that test subjects do for presenting their biometric characteristics to the biometric capture device and during the complete interaction with the biometric system. For most of the cases, the use of a multi-camera recording system is recommended, in order to obtain different views of the user interaction.

6.4 Fundamental requirements for planning a H-B interaction testing of biometric systems

As it was described in section 6.2.5, H-B interaction testing involves a biometric performance evaluation. For the proposed methodology the most proper evaluation type is scenario evaluation. A scenario evaluation obtains biometric performance of a complete biometric system testing under controlled conditions which model the real application and its target population (see section 3.5.1.2).

This section specifies all essential requirements for planning the H-B interaction testing of biometric systems in compliance to a biometric performance scenario evaluation addressed by ISO/IEC 19795 Part 1 and 2. Basically, it establishes a generic scenario evaluation which has been adapted to analyse the influence of H-B interaction factors.

Figure 16 shows all aspects that must be addressed and which of them have been modified for human-biometric interaction testing. The latter have been indicated in blue. Alternatively, some aspects shall be just modified when the purpose of the evaluation will be to analyse any of the factors which has been described in section 6.3.1. These aspects have been identified using the green colour for factors that belong to the biometric capture device group, red for factors that belong to the human being group and purple colour for factors that belong to the interaction group. Later in this chapter all of them have been described in order to provide a complete methodology although some of them do not need any modification[3].

[3] Those aspects that are basic for a scenario evaluation (i.e. the aspects coloured in white in the figure) have been defined in a similar way to the environmental testing methodology. As a consequence, their corresponding sections have identical text to Chapter 5. Nevertheless, these have repeated for preserving the independence of both methodologies.

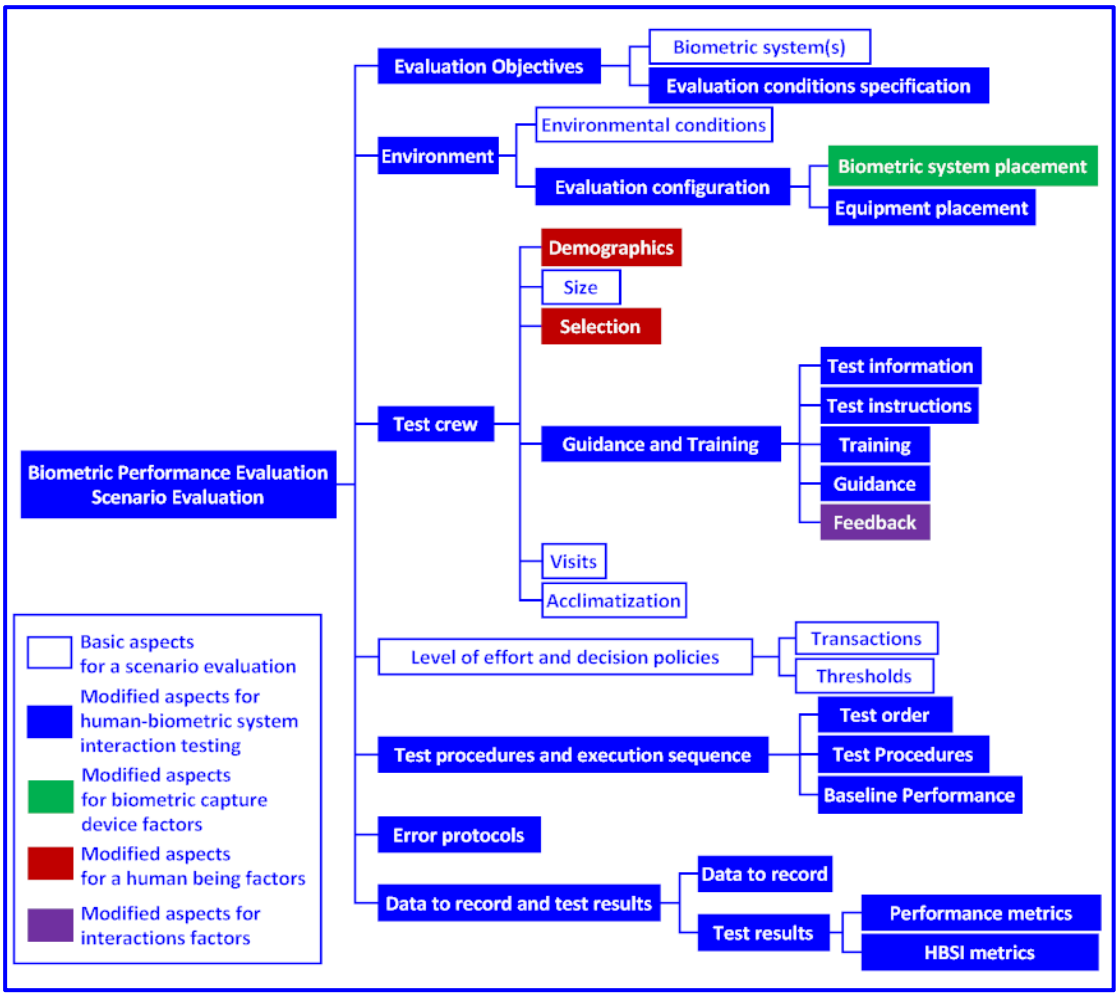

Figure 16. Scenario evaluation specification according to ISO/IEC 19795 Part 1 and 2 for H-B interaction testing

As it will be explained below, most aspects are dependent of the intended application and the target population and shall be specified by the parties involved in the evaluation according to the evaluation objectives. In addition, other aspects shall be defined per each evaluation condition, so it is required that the test plan covers both the REC and the TEC(s).

6.4.1 Define evaluation objectives

For a scenario evaluation, the first step is to define the objectives of the evaluation. These shall include the following:

- A description of the biometric system(s) under test. This consists of an explanation of the biometric system(s), its modality, its capture device(s), as well as the main components that compose it. Also, it shall be described if the recognition process is based on verification (one-to-one) or identification (one-to-many) functions. For the latter, it shall be specified if it is an open-set identification or a closed-set identification too.

- A guide of the biometric system functionality. This guide must include a description of biometric functions which are implemented in the biometric system, how these functions work and their input and output parameters. This guide will be used for defining some requirements for the scenario evaluation.
- A description of the expected application including the intended operational environment (either for enrolment and recognition) as well as the target population. If it is unknown or the H-B interaction testing is independent of them (i.e. the intention of testing is analyse a predefined specification), it shall be clarified.
- The objective of H-B interaction testing: to analyse whether one or a set of H-B interaction factors can affect biometric system performance, or not, and quantify their influential effects or to obtain biometric system performance considering a particular H-B interaction factor specification in comparison to a variation to such specification.
- The evaluation conditions specification. A statement that claims the factor(s) and its specific variation(s) to assess. It shall be specified in compliance to section 6.3.
- The definition of the REC and TEC(s) to test in accordance to the evaluation conditions specifications mentioned in the previous bullet. Each evaluation condition shall be described detailing the following:
    - Type of the evaluation condition: reference or target.
    - Evaluation conditions specification for enrolment including the factors' specification as well as the necessary equipment and instructions for generating, controlling and recording such specification.
    - Evaluation conditions specification for recognition including the factors' specification as well as the necessary equipment and instructions for generating, controlling and recording such specification.

6.4.2 Operational environment

All scenario evaluation shall be carried out in an operational environment. In order to specify the scenario evaluation two aspects have to be defined: the environmental conditions and the evaluation configuration. As in this case the influence of the ambient conditions are not going to be evaluated, the environmental conditions will be kept constant throughout the evaluation (both at REC and at TEC), and shall be planned in a similar way that any scenario evaluation, using defined values.

However, in an H-B interaction testing process, the evaluation configuration will depend on the particular evaluation conditions to be tested. The following subsections address requirements to plan both of them.

6.4.2.1 Environmental conditions

The ambient conditions for each scenario evaluation shall be similar to the intended environment. If this environment is unknown, these values shall be selected considering any of

these possibilities: conventional standard conditions (See Table 20 and Figure 17), a reasonable range in accordance to the biometric system specification or the typical values of the test laboratory (considering indoor conditions).

Table 20. Standard conditions for the environmental parameters

Environmental parameter	Standard conditions value
Temperature	23 °C (± 3 °C)
Relative humidity	40% to 60% (± 5%)
Illumination	Fluorescent light - Colour temperature: 3300K to 5300K Illuminance: 300 lx to 1500 lx (± 5%) Irradiance: typical spectrum for fluorescent lamps
Noise	$L_{p,A,eq,T}$ < 65 dB(A) (±3 dB) being T= time for a user biometric transaction $L_{p,Cpeak}$ < 70 dB(C)

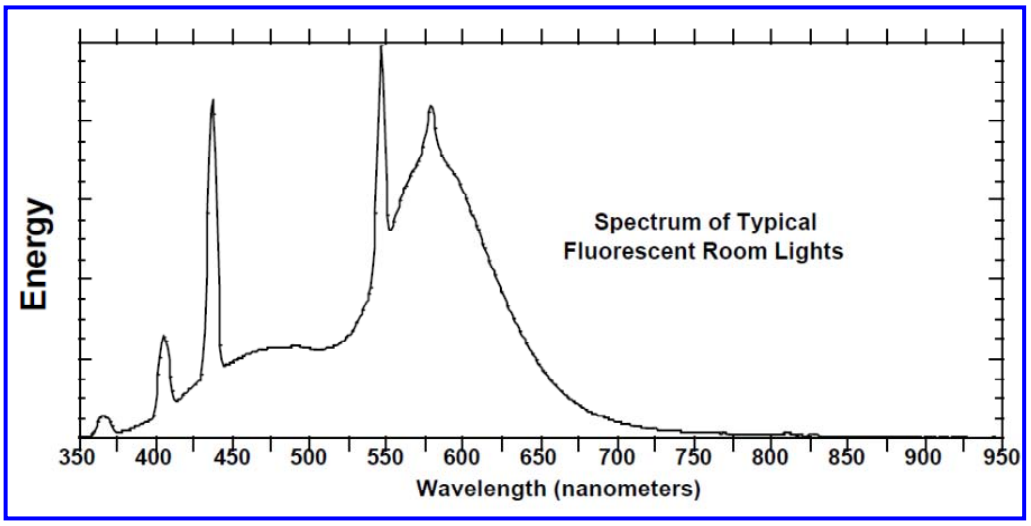

Figure 17. Spectrum of typical fluorescent lamps [ASD'99]

Nevertheless, whatever values are defined it is indispensable that the test laboratory is able to reach them without any additional equipment or with equipment that do not interfere in test subjects interactions. It is also a requirement to be able to keep such conditions within the pre-defined ranges during the whole evaluation process.

6.4.2.2 Evaluation configuration

The operational environment also shall be specified in terms of where the biometric system(s) and the necessary equipment are located. For planning both issues, the following requirements shall be met.

6.4.2.2.1 Biometric system(s) placement

When the factors to assess involve the biometric capture device(s) position, the biometric system(s) may be located for each evaluation condition in such way that the biometric capture device(s) meets the corresponding specification defined per each evaluation process (i.e. enrolment or recognition) that is going to be tested at every moment. For doing that,

requirements addressed in section 6.3.5 to generate such evaluation conditions shall be satisfied.

Otherwise, the biometric system under evaluation should be located in the specified evaluation configuration in a consistent manner with the target application or biometric system supplier's recommendations. But this manner must allow test subjects to interact easily. In any case, biometric system placement shall be selected in agreement among all parties involved in the evaluation.

6.4.2.2.2 Equipment placement

As it was described in section 6.3.6, for H-B interaction testing is an indispensable requirement to record the test subject's behaviour during their interactions with the biometric system(s). Therefore, the test plan shall include the location of video camera(s) in the evaluation scenario. Such locations shall be chosen by the parties involved in the evaluation following the requirements given, in order to record test subjects' behaviour, but not disturbing their interactions.

6.4.3 Test crew

The set of test subjects that are going to participate in a scenario testing is called test crew. It has been demonstrated that the characteristics of the test crew influence on biometric performance [DOD'98]. Therefore, people that take part in the evaluation (i.e. the test subjects) shall fulfil the requirements given in the following subsections.

6.4.3.1 Test crew demographic characteristics

Test subjects shall be people which have representative characteristics of the target users. That is, test crew shall be composed by a percentage of people whose gender, age, ethnic origin and occupation or technical knowledge will be similar to the final target population.

When the factors to assess entails a specification of characteristics or elements that test subjects must have and it is not possible to provide them at the testing facilities, the people that compose the test crew shall fulfil such characteristics. In this case, it could be difficult to satisfy the representativeness of the target population requirement at the same time of the specification of the factor. In such case, the fulfilment of the specification of the factor(s) shall take precedence.

6.4.3.2 Test crew size

The number of test subjects that make up the test crew shall be large enough to achieve statistically significant results. The ISO/IEC 19795-1 standard establishes the 'Rule of 3' or 'Rule of 30' to calculate the number of recognition attempts that is necessary to carry out for obtaining results at specific confidence levels. Based on this number and considering other related factors like the number of visits, the number of attempts carried out per each test

subject, the availability of resources and cost and time constraints, parties involved in the evaluation shall determine the test crew size.

Due to the fact that some test subjects will probably leave the evaluation at any stage, not completing all programmed visits, it is recommended to increase test crew size in around a 10%.

For testing biometric systems based on open-set identification functions, it will be indispensable to have a group of test subjects who will not be enrolled for conducting impostor transactions. This special group shall fulfil the same requirements addressed for the common test subjects excluding those requirements related to enrolment.

6.4.3.3 Selection of test subjects

The selection of test subjects shall be random in terms of not allowing to recruit test subjects for whom the ability to recognize them is previously known. Nevertheless, the selection process shall conform to demographic requirements given in section 6.4.3.1, especially when a factor that must have test subjects is analysed.

Moreover, test subjects must not have been involved in design, development and implementation processes of the biometric system under test and/or must not have been participated in recognition algorithm training or tuning procedures.

6.4.3.4 Guidance and training of test subjects

Another relevant factor of the test crew which influence on biometric performance is the way that they carry out their interactions and the different level of habituation of test subjects. Depending on guidance and training procedures, test subjects can improve their interactions and the level of habituation can be balanced among test subjects reducing its influential effects significantly. Considering this situation together with the factors that are going to be tested, test subjects shall be informed, guided and trained according to the following requirements.

In case that multiple biometric systems are going to be assessed, instructions, guidance and training shall be planned considering all of them.

6.4.3.4.1 Test information

Test subjects shall be informed about the evaluation process including an overview of the evaluation, its purpose, the number of times that they must attend the testing facility, the duration of each visit and other relevant information such as legal issues related to data protection or privacy policies.

Regarding the H-B interaction testing, people shall be informed about the evaluation conditions; especially if there is any element that it is indispensable for the evaluation process

that test subjects shall wear when they have to come for participating in the experiments such as contact lens, piercings and so on.

It is suggested to develop forms which include the complete information about the evaluation and a declaration of acceptance to participate in it. These forms shall be signed by users before turning into test subjects.

6.4.3.4.2 Test instructions

Once people have been designated as test subject, they shall be informed about the evaluation steps and what they have to do at each step. This explanation shall be developed according to the target application and have to include the following information:

- A description of enrolment and recognition functions, how to execute them, the number of attempts, which data must be provided by test subjects and which information are the test subjects going to receive from the biometric system.
- Instructions about how to provide the biometric characteristic to the capture device considering right and non recommended actions as well as possible information given by this device. For assessing certain factors, these instructions must include certain description according to the factor's specifications.
    - When assessing factors that depend on the biometric capture device, these instructions shall include a recommendation about that test subjects shall present his biometric characteristic to the device in a consistent manner regarding the position of the biometric capture device.
    - When assessing factors that depend on the human beings, these instructions shall address the following considering the specific case:
        - Physical elements: elements that test subjects must wear when they are going to interact with the biometric system.
        - Emotions: expressions that test subjects must do including the level of expressiveness when they are going to interact with the biometric system.
        - Chemical products: products that test subjects must apply before they interact with the biometric capture device and how to apply it or in what to extent.
    - When assessing factors that involve translations and rotations for the presentation of the biometric characteristic, these instructions shall explain how to perform these particular interactions apart to the correct way above mentioned. For this case, instructions about each type of interaction shall be provided to test subjects before they have to perform the corresponding type in order to avoid causing confusion to them. Nevertheless, it will explain in sections 6.4.5.2 and 6.5.

6.4.3.4.3 Training

Before the beginning of tests in every evaluation conditions, test subjects shall perform practical enrolment and recognition attempts under such evaluation conditions. The requirements for these practical attempts are that test subjects shall interact according to factor's specification regarding to the position of the biometric capture device, the necessary elements that they must wear or presenting their biometric characteristic as it has been defined for such evaluation conditions respectively. During these attempts, test operators shall supervise test subjects actions and correct any mistake considering the particular circumstances of each kind of evaluation condition. This training phase shall be adapted to the skills of each test subject and it must last till test subjects demonstrate proficiency in their interactions with the biometric system.

6.4.3.4.4 Guidance

Test subjects shall be guided during training. During enrolment and recognition it depends on the target application and the objectives of the evaluation, so it shall be decided by parties involved in the evaluation. It is recommended to guide both processes if they are controlled processes subjected to supervision or attended processes. Otherwise, enrolment and recognition should not be guided.

Nevertheless, although enrolment and recognition are decided to be non-guided processes, both shall be supervised by test operators. Such test operators shall intervene at any moment if they observe certain errors. The specific errors and the related actions to perform will be described in section 6.4.6.

In any case, guidance shall be defined during the evaluation planning in a consistent manner to test instructions including points in which guidance is required, localization of test operators to provide them, and the specific guidelines that test operator shall give to test subjects. For H-B interaction testing, such guidelines shall be adapted to the particular evaluation conditions as necessary. Therefore, in a similar way to test instructions, it may be needed to develop specific guidelines for each factor specification to be tested.

6.4.3.4.5 Feedback

The last factor regarding training and guidance is the feedback. Feedback refers to the information about the process which is provided to users by the biometric system and/or the biometric capture device by means of a display, lights or sounds.

When testing the H-B interaction factors that depend on the biometric capture device or human beings that are covered by the current proposed methodology, there is not any specific requirement it considering just factors. Just, if the biometric system and/or its capture device provide any kind of feedback to users, it shall be given to test subjects for improving their interactions in a similar way to the final application.

However, when testing factors in which test subjects shall modify their interactions deliberately, the feedback provided by the biometric system may not be appropriate. In these

cases, partied involved in the evaluation shall decide whether any kind of feedback is provided (including the most appropriate type) or not for each evaluation condition.

6.4.3.5 Visits

Visit is a concept that refers to each time that test subjects must attend to the test laboratory for carrying out evaluation activities. Regarding this aspect, ISO/IEC 19795 Part 1 and Part 2 addresses the following:

- Multiple visits allow increasing the number of recognition transactions for only a slight rise of the evaluation cost. It is easier to get that test subjects come back to the test laboratory than to recruit new test subjects.
- Several visits allows to observe the influence of factors related to users on biometric performance such as the level of habituation (which usually improves biometric performance) or template ageing (which typically gets worse performance).
- There shall be a time separation between enrolment and recognition attempts. .

Considering these circumstances, evaluations shall have more than one visit. These visits shall take place at different times. The separation interval shall be defined in compliance to the separation time between enrolment and recognition processed at the target application.

6.4.3.6 Acclimatization

Acclimatization refers to the time that takes the human body to adapt to certain environmental conditions. If test subjects are not acclimatized, it may affect to the biometric sample acquisition process in a greater extent that the H-B interaction factor under test.

Therefore, acclimatization procedures should be established as necessary. Each procedure shall include the following:

- times in which this approach shall be carried out,
- minimum duration of the period for acclimatization,
- mechanisms and test subject actions to achieve acclimatization, and
- criteria to consider that test subjects are acclimatized.

It is important to consider the time that takes this process when planning the evaluation. This time may increase the duration of tests and, as a consequence, it might cause tiredness and a lack of motivation in test subjects.

6.4.4 Level of effort and decision policies

Other relevant factor of a scenario evaluation is the specification of the number of times that test subjects have to interact with the biometric system and the constraints of these interactions. This aspect is referred as level of effort and decision policies and shall meet the

same requirements established for a regular scenario evaluation. Once this has been specified it will be similar for all evaluation conditions.

6.4.4.1 Transactions

In order to obtain performance rates, test subjects shall be enrolled and shall execute recognition transactions. These transactions shall be as follows.

- Enrolment transactions are for generating biometric references of the test subjects. So, all test subjects shall execute this type of transaction once at each enrolment evaluation conditions except for biometric systems which operation mode is an open-set identification. For those systems the special group of test subjects selected for impostor transactions must not be enrolled. Depending on the expected evaluation effort and the biometric modality such enrolment may generate various biometric references. Each of these shall be correctly identified in order to avoid errors.
- Recognition transactions are for checking biometric recognition functions. These transactions shall be verification transactions for testing biometric systems based on verification functions and identification transactions for testing those systems based on identification functions. In any case, test subjects shall carry out two different types of recognition transactions: genuine and impostor transactions.
    - Genuine transactions. For these transactions the test subject shall be previously enrolled at the system and it shall provide his own biometric characteristic. When testing a biometric system based on verification functions, the test subjects shall provide their own identifier as well. Such identifier shall be the correct one in order to avoid errors. In case of close-set identification functions, either the test subject or the test operator shall confirm whether the identified user corresponds to the test subject. In both cases the complete test crew shall execute this type of transactions.

      On the other hand, when biometric systems based on open-set identification functions are tested, genuine transactions shall be only executed by common test subjects providing just their biometric characteristic. The special group designated for performing impostor transactions, as it has not been enrolled, is expected to provide a recognition error in their genuine transactions.
    - Impostor transactions. For performing these transactions test subjects shall provide their own biometric characteristics.

      When analysing biometric system based on verification functions, all test subjects shall execute impostor transactions. In addition to their biometric characteristic, either the test subjects, the operator, or the evaluation system (e.g. chosen randomly) must provide the identifier of other enrolled test subject. Such identifier shall be selected randomly from available templates but excluding of the candidates those identifiers that

belong to templates of the particular test subject who is going to execute the impostor transaction. This is a must because it is not a good practice to conduct impostor transactions in which samples of the same test subject are compared.

When analysing biometric systems based on open-set identification functions, only the special group of test subjects shall execute impostor transactions. In this case, test subjects do not have to provide any kind of identifier.

At last, when analysing biometric system based on closed-set identification functions, this type of transactions shall not be executed.

Furthermore, it shall be specified the number of recognition transactions that each test subject must carry out per visit. This number shall be determined together with the number of visits and the test crew size, as a result of applying the 'Rule of 3' or 'Rule of 30', as it was explained in section 6.4.3.2. It is important to note that both rules are dependent of the expected error rates, so the number of genuine transactions may be different to the number of impostor transactions.

Moreover, a transaction may consist of one or more number of attempts and each attempt may consist of certain number of presentations. Therefore, the maximum number of presentations per attempt and attempts per transaction shall be specified. In addition, presentations, attempts and transactions may have a limited time to be executed. Therefore, the maximum time for accomplishing a presentation, attempt and/or transaction shall be defined as well. All these settings shall be consistent with the target application.

When testing several biometric systems, it shall be decided if the number of presentations/attempts/transactions will be identical across all systems or change according to the operation of each system. This decision concerns to parties involved in the evaluation who shall assess possible effects to modify the number of presentations/attempts/transactions for biometric systems under test or the difficulty to deal with different numbers during the evaluation process.

As a general requirement, all attempts (and transactions) shall be done with disengagement from the device. In other words, test subjects shall execute the action to present their biometric characteristic to the capture device and then the action to remove the biometric characteristic from it per each attempt. It is not allowed that test subjects present their biometric characteristic to the capture device once and keep it positioned there to carry out all attempts.

6.4.4.2 Thresholds

Some biometric systems have configuration options that let customers to select quality and decision thresholds. When it happens, these parameters shall be fixed in a consistent manner with the target application. If quality thresholds are different for enrolment and

recognition processes, the corresponding parameter for each process shall be identified and reported.

6.4.5 Test procedures and execution sequence

After establishing the requirements for all elements that are involved in the evaluation, i.e. environment, test crew and biometric system, specific procedures shall be planned for conducting the scenario evaluation in each evaluation condition. Such test plan shall satisfy the following requirements.

6.4.5.1 Testing order of evaluation conditions

The order of testing evaluation environments shall be random with the intention that effects like habituation or test subjects tiredness affects biometric performance as less as possible.

However, H-B interaction testing requires conducting two scenario evaluations at least: one for the REC and another for the TEC. As the number of factors to analyse will be higher, the number of TECs and the scenario evaluations to carry out will be also higher. As a result, the time and the effort needed for the evaluation will increase significantly. Considering these circumstances, a reasonable order of the evaluation conditions to test may help to reduce them.

For this reason and when there are multiple TECs to analyse, it is allowed to apply semi-randomness in the order. This fact shall be justified properly. Reasons for a semi-random order could be:

- to minimize the time to change the evaluation configuration
- to minimize the time of training test subjects, or
- for the availability of equipments.

When H-B interaction testing entails the evaluation of several biometric systems, the order of executing test subjects interactions in each system under the same evaluation conditions shall be random too.

6.4.5.2 Test procedures and its execution sequence in terms of visits

In addition to establish a test order for the evaluation conditions, it is necessary to plan the overall evaluation. Specifically, the plan shall include visits and which tasks to be executed in each visit by test subjects.

According to requirements already stated, at the first visit test subjects shall perform training and enrolment. Only for biometric systems based on verification functions it would be possible to carry out the first session of genuine recognition transactions at those visits. At subsequent visits, test subjects shall just perform different sessions of recognition transactions in all the evaluation conditions. When testing factors that correspond to the interaction

process, training shall be conducted before the test subjects carry out their interactions in each evaluation condition. Therefore, for this case training shall also be conducted at the subsequent visits.

In general, it is suggested to develop flowcharts which include the people and the roles taking part in each test activity (i.e. test operators, test subjects, etc).

Within the test procedures planning, it shall be also decided how to arrange test subjects visits. Test subjects may come to the test laboratory alone or in a group. For the former situation, evaluation conditions are changed per each test subject whereas for the latter situation, all test subjects will carry out their recognition transactions before changing the evaluation conditions. Again, this aspect shall be determined by parties involved in the evaluation in a consistent manner with the difficulty to install and change the configuration of the evaluation conditions, the availability of test subjects and other factors that may modify the duration of the visits like training or acclimatization.

6.4.5.3 Establishment of a baseline performance

Regarding test procedures, there is another aspect that must be considered for H-B interaction testing. This is the establishment of a baseline performance. That is, the specific procedures for obtaining reference results at predefined reference specifications. In general and according to the evaluation model, these procedures consist of carrying out the defined scenario evaluation at REC.

For the H-B interaction factors that cover the proposed methodology, the baseline performance shall be obtained following the general requirement, i.e. carrying out the specified scenario evaluation at REC for the reference specification.

Then, for quantifying the influence of H-B interaction factors, results for the target specification shall be compared against the results obtained for the reference specification. This will be further explained in section 6.4.7.2.

6.4.6 Error protocols

During the evaluation, different errors can occur. The test plan has to specify actions that test operators shall accomplish to assure that errors do not affect evaluation results. Depending on the kind of errors, these actions shall be the following:

- General errors: these errors happen when the biometric capture device does not work correctly. In this case, the test operator shall stop the evaluation and solve the problem. Once the biometric system works properly again, the evaluation can continue.
- H-B interaction anomalies: if test operators detect that any of the requirements for the factor's specification under testing has not being fulfilled, they shall stop

the evaluation and correct the possible mistake. Once it has been solved, the evaluation can resume.

- Enrolment and verification errors: if test operators detect that the test subject has introduced a wrong identifier or has presented a wrong biometric characteristic, they shall cancel the attempt/transaction, inform the test subject about the error and the particular attempt/transaction shall be repeated by the test subject.

6.4.7 Data to record and test results

The last aspect that shall be planned for the environmental testing evaluation is the information to be recorded during experiments and how to calculate test results. If the necessary data to quantify biometric performance is not saved, it will be not possible to obtain evaluation results. As a consequence, the effort dedicated to the evaluation will be in vain.

6.4.7.1 Requirements for recording data

Fundamental data that shall be recorded for each evaluation condition must be the following:

- the outcome of the biometric enrolment or recognition attempt/transaction,
- videos about test subject interactions,
- all kind of errors, and
- any essential information for obtaining the mandatory results addressed in the next section.

It is suggested to save as much information as possible related to the outcome of the biometric enrolment and recognition attempts/transaction. It makes a broader analysis of the evaluation results possible. Next, recommended data to save are specified. It is important to note that it will be not always possible to record the complete list of the below mentioned data.

- For an enrolment attempt/transaction:
  - Test subject demographics characteristics who executed the attempt/transaction.
  - Biometric characteristic(s) which are enrolled.
  - Identifiers assigned to the test subjects.
  - Results of the enrolment process (successful/Failed).
  - Number of presentation/attempts needed.
  - If enrolment fails, the possible cause.
  - Quality score of the biometric sample.
  - Date and time when the attempt/transaction is executed.
  - Duration time of attempt/transaction.
  - Other relevant data (e.g. settings for the enrolment such as quality and decision thresholds).

- For a recognition attempt/transaction:
  - Test subject identifier.
  - Type of attempt/transaction: genuine or impostor.
  - Biometric characteristic which is used.
  - For impostor attempt/transaction, the identifier of the test subject who presents his biometric characteristic.
  - Similarity score or successful /failed recognition or candidate list.
  - Number of attempts needed.
  - If biometric capture or acquisition process fails, the possible cause.
  - Quality score of the biometric sample.
  - Date and time when the attempt/transaction is executed.
  - Duration time of attempt/transaction.
  - Other relevant data (e.g. settings for the recognition process, such as quality and decision thresholds and/or the number of identifiers to include at the candidate list).

If in addition to these data, biometric samples are saved, it will be also possible to do *offline* testing although this kind of testing is not able to reflect all the H-B interaction influential effects as it was described in section 6.2.

Moreover, due to the significant amount of data generated during tests, it is recommended to automate the process as much as possible. With automated tools and processes test operator‘s work becomes easier and it prevents from human errors. Evaluation ends up being more independent and reports will be generated more easily.

6.4.7.2 Test results

Once tests have been finished, biometric performance results shall be calculated for each factor and its corresponding specification system under test. Specifically, these results shall consist, at least, of the following measurements:

- Performance metrics including error rates and throughput rates:
  - Acquisition and signal processing:
    - Enrolment: FTE rate, the minimum, maximum, arithmetic mean and standard deviation time that takes to carry out an enrolment attempt/transaction.
    - Recognition: FTA rate, the minimum, maximum, arithmetic mean and standard deviation time that takes to acquire the biometric sample.
  - Comparison and decision processes:
    - Only for biometric systems based on verification functions:
      - FNMR and FMR rates. These rates may be given using ROC and/or DET curves.

- The minimum, maximum, arithmetic mean and standard deviation time that takes a comparison attempt.

- Complete recognition process:
  - For biometric systems based on verification functions:
    - FRR/FAR and GFRR/GFAR rates. These rates may be given using ROC and/or DET curves.
    - The minimum, maximum, arithmetic mean and standard deviation time that takes a verification transaction.
  - For biometric systems based on open-set identification functions:
    - FNIR and FPIR rates. These rates may be given using ROC and/or DET curves.
    - Identification rate. For multiple ranks, this rate may be given by means of CMC curve.
    - The minimum, maximum, arithmetic mean and standard deviation time that takes an identification transaction.
  - For biometric systems based on closed-set identification functions:
    - FNIR rate.
    - Identification rate. For multiple ranks, this rate may be given by means of CMC curve.
    - The minimum, maximum, arithmetic mean and standard deviation time that takes an identification transaction.

In addition, all measurements shall be given together with the number of attempt/transactions used to obtain these measurements and their uncertainty. In case of biometric system based on identification functions, the number of templates that takes part in the comparison process shall be provided.

- Human-Biometric Sensor Interaction metrics. Regarding these metrics it is required to calculate the HBSI metrics that correspond to the segmentation of the FTA. That involves the following metrics.
  - Erroneous presentation:
    - Defective Interaction rate (DI)
    - Concealed Interaction rate (CI)
    - False Interaction rate (FI)
  - Correct presentation:
    - Failure to Detect rate (FTD)
    - Failure to Process rate (FTP)
    - Successful Processed Sample rate (SPS)

For obtaining them, two steps shall be carried out. On one hand, the test plan shall include the definition of test subject's interactions in terms of tasks as well as the test subject's actions and events that may happen in each task. Besides, the different actions and events shall be associated to a correct/erroneous presentation and the type of HBSI metric. As it was explained in section 6.2.3, this

specification of shall be defined by the parties involved in the evaluation considering the biometric system under test, its modality, the biometric capture device and the target application. On the other hand and once tests' subjects have finished their participation in the evaluation, it will be necessary to analyse the video recordings that contains test subjects' interactions and classify them according to first, the type of presentation and then, the type of metrics.

Regarding the rest of the H-B interaction metrics mentioned in section 6.2.3, it is recommended to obtain them, but it is not required. The reason is because these metrics do not measure the overall influence on biometric system performance.

Once results have been obtained for each evaluation condition, results shall be calculated for the H-B interaction testing evaluation. Such results disclose the H-B interaction factor influence on biometric performance. For this purpose, each performance metric (referred as "X") shall be generated from the comparison of the target specification evaluation results against the baseline performance results as it is expressed in the following equation:

$$X_{\text{Factor specification influence}} = X_{\text{Target}} - X_{\text{Baseline}} \quad (10)$$

Furthermore, it is also necessary to offer additional information about the overall evaluation process such as:

- Test crew demographics composition.
- A distribution time between visits.
- Error logs and general observations about the complete evaluation process.

6.5 Fundamental requirements for executing a H-B interaction testing of biometric systems

Once the test plan has been developed, the next step is to conduct H-B interaction testing in compliance with such plan. A consistent set of sequential activities shall be executed by test operators and test subjects for each of the evaluation conditions. These activities have been detailed in the next subsections[4]. When the group of activities are not listed in order, it is because the order is not relevant.

6.5.1 Pre-test activities

The test laboratory shall conduct several actions prior to conduct the evaluation conditions experiments. These shall be the following:

- Examine the biometric system(s) under test and implement the essential testing support application for performing the evaluation. It shall be able to collect the

[4] These subsections contain quite similar requirements to the environmental testing methodology but adapted to the H-B interaction testing methodology.

specified information and shall be conformant with the levels of effort and decision policies defined.

- Develop a plan for recruiting the needed test subjects and how these people are going to be identified.
- Develop a general evaluation schedule for arranging test subjects visits.
- Implement evaluation acceptance forms, data forms and guidelines for test subjects.
- Instruct test operators about how the biometric system works, how to use the evaluation application, how to handle equipments, how to guide and train test subjects and all necessary details to carry out the evaluation
- Develop check lists and forms which allow test operators to detect and write down errors.
- Select the necessary equipment for recording test subjects interaction, check their correct operation and verify the corresponding methods for saving the essential information.
- Prepare the lay out for the biometric system and recording devices. It may entail to make a particular structure to locate them.
- Prepare additional resources for the evaluation (e.g. devices for accomplishing acclimatization procedures, tools for installing the evaluation configuration, elements or products essential for the evaluation, etc).

In addition, it is recommended to perform a mock evaluation in which one test operator has a test subject role in order to detect if something is missing or in order to check how long it takes. Sometimes, from the results obtained in this mock test, it might be needed to modify the test plan.

6.5.2 Test activities

Once, everything is ready for the evaluation, test subjects interactions shall be executed in each evaluation condition. For this purpose, the following actions described in the following subsections shall be carried out.

6.5.2.1 Procedures before the first visit

At the very beginning, some tasks shall be completed before the test subjects interactions. These are the following:

- Recruit test subjects giving them appointments to come to the test laboratory at least for the first visit.
- Install the evaluation configuration in which test subjects shall execute their training including biometric system(s) and equipments.
- Verify the correct operation of biometric system covering all biometric functions that is going to be tested.

6.5.2.2 First visit

During the first visit, test operators and test subjects shall execute multiple tasks in the following order:

1. Test operators shall explain test information to test subjects and test subjects shall fill in evaluation acceptance forms.
2. Test operators shall explain test subject instructions to test subjects.
3. Test subjects shall carry out practical trials at the evaluation configuration till they demonstrate proficiency in biometric system interactions.
4. When the training will be finished, test operators shall prepare the enrolment evaluation conditions and check that all, biometric system(s), equipments and the evaluation application for recording data work satisfactory.
5. Test subjects shall execute enrolment process. If acclimatization procedures are necessary, these shall be done before test subject interactions begin. Test operators shall guide this process in accordance with the test plan. They also shall solve any error that occurs and write it down on the error logs.
6. Dismantle the evaluation conditions as necessary depending on the next steps of the evaluation.
7. If test subjects shall carry out enrolment in further factor's specifications, the steps 4 to 6 shall be repeated for the rest of evaluation conditions. The order shall conform to the test order established at the evaluation plan.
8. The subsequent visits shall be set if it was not done previously.
9. Test operators shall save all data collected during this visit in a safe way.

In case of testing biometric systems based on verification functions, the steps 2 to 4 described in the next section could be carried out at the first visit but only for genuine recognition transactions.

6.5.2.3 Subsequent visits

For the rest of visits, test operators and test subjects shall carry out similar tasks to the first visit excluding those tasks related to enrolment. Specifically, the order for tasks shall be the following:

1. Test operators shall remind briefly test instructions to test subjects. At least the tasks to conduct during this kind of visits.
2. Then, the first recognition evaluation condition shall be installed by test operators. Again, they shall check that all devices (i.e. biometric system(s), equipments and the evaluation application for recording data) work properly.
3. Test operators shall assure that the corresponding factor's specification for this evaluation condition is met.
4. Test subjects shall carry out practical trials at the evaluation configuration if it is different to the last evaluation configuration in which they have taken part.

5. Test subjects shall execute the first session of recognition attempts/transactions in the evaluation conditions. It entails either genuine and impostor attempts/transactions. Test operators shall guide this process in compliance to the test plan. They also shall solve and write down any inconvenience that occurs.
   In case of impostor transactions for a biometric system based on verification functions, test operators shall provide the test subject with the identifier of the template which will be forged.
6. Dismantle the evaluation conditions as necessary depending on the next steps of the evaluation.
7. Steps 2 to 6 shall be repeated for all the recognition evaluation conditions to test following the order established at the test plan.
8. Then, test operators shall save all data generated during the visit in a safe way.

6.5.3 Post-test activities

Finally, test operators shall calculate results and develop the corresponding reports. In particular, they shall perform the following actions.

- Obtain results per each evaluation conditions.
- Calculate the general results for the H-B interaction testing evaluation comparing results from the TEC to baseline results.
- Obtain conclusions. It is recommended to analyse error logs, video recordings and any relevant information for doing this task.
- Generate the evaluation report. This report shall include all the information stated in the next section.
- Close the evaluation. It may entail tasks such as storing all relevant information according to the test laboratory policies; remove personal data in compliance to data protection laws, dismantle biometric system(s) and other equipment, etc.

6.6 Fundamental requirements for reporting a H-B interaction testing of biometric systems

As it has been mentioned in the previous section, the last part of the evaluation is to develop a report which gathers the results and the test procedures used for obtaining them. This report shall include the information specified as follows[5].

- The test plan. This document shall include all aspects that have been defined in section 6.4 as mandatory aspects to be specified either for the scenario evaluation or for H-B interaction testing.

[5] This section contains quite similar requirements to the environmental testing methodology but adapted to the H-B interaction testing methodology.

- Any modification performed to such test plan. This modification shall be described and justified.
- Final size of the test crew and its composition.
- Distribution time of test subject visits and how many test subjects have participated in each visit.
- For each evaluation condition:
    - A description of the particular factors specification that has been tested.
    - The specific evaluation configuration used for this evaluation condition by means of photographs or diagrams. It shall include the location of equipments for recording test subjects' interactions.
    - Test results addressed in section 6.4.7.2.
    - Errors that have occurred during the experiments for this evaluation conditions.
    - Any relevant comment considering error logs for the obtained results.
- The baseline performance results shall be indicated clearly.
- General results of the H-B interaction evaluation as well as an analysis which interprets them. It is recommended to provide graphics which include similar measurements at different evaluation conditions. These graphics are very helpful when analysing results.
- Final conclusions for the overall evaluation.

6.7 Experiments developed for validating the methodology

Once the whole methodology has been explained, this section describes different experiments that have been conducted for developing, improving and validating the proposed H-B interaction testing methodology for biometric systems. This description has been divided in two sections. The first section describes the preliminary studies that were carried out and the first version of the methodology. Then, the second section explains the evolution of this methodology highlighting those points which were improved and the future improvements to the proposed methodology.

6.7.1 Preliminary studies and first version of the evaluation methodology

The motivation for the development of the H-B interaction testing methodology came from the development of the environmental testing explained in Chapter 5. During the analysis of the influence of the environment on biometric performance it was noted that both, ergonomics and the users' behaviour are other important aspects that may affect biometric system performance. In fact, human factors and its behaviour may diminish biometric performance in a greater extent that environmental conditions.

Therefore, a preliminary research work was conducted for analysing the existing works on this area. However, as it has been explained in section 6.1, although there were several studies about this topic, no formal evaluation methodology to analyse and quantify the user's

behaviour existed. As a result, the first version of this methodology was proposed. This can be seen in the article title "Evaluation Methodology for Analyzing Usability Factors in Biometrics"[FER'10b].

In particular, this work shows an initial classification of the influential factors and the specification of the evaluation methodology to analyse some of them. Moreover, the evaluation methodology was applied to a fingerprint verification system for checking its viability. Five evaluation conditions to assess were specified, expressed as different "scenarios". These have been summarized in Table 21.

In this case, the "scenario 1" entailed the REC. These conditions were called as "common scenario" and were also used in the rest of "scenarios" for those aspects which were not the target of the evaluation in each respective TEC. Enrolment was also carried out once, considering its conditions as the same ones as the "common scenario".

Table 21. Evaluation conditions for recognition [FER'10b]

Evaluation conditions	Type of evaluation conditions	Factor	Possible variations		Definition
Scenario 1	Reference	----	---	---	Common Scenario
Scenario 2	Target	Position	Inclination		Slope of 40º
Scenario 3	Target	Human-biometric capture device interaction	Presentation of the biometric characteristic	Translation	1 cm up from the centre of the usable fingerprint area
Scenario 4	Target	Human-biometric capture device interaction	Presentation of the biometric characteristic	Translation	1 cm down from the centre of the usable fingerprint area
Scenario 5	Target	Biometric characteristic	Temporary conditions (It can be removed for the interaction)	Chemical products	Hand cream applied to the complete hand

This definition of the "common scenario" and the overall evaluation entailed the following test plan. The test plan was explained emphasizing those cases in which it was necessary to modify the conditions for fulfilling each of the aforementioned "scenarios".

- Environment. The environmental conditions were typical indoor conditions with a temperature of 22±4°C and a relative humidity between 40 - 60 %. Illumination was fluorescent light with an intensity varied between 1,500 and 2,200 lx in the visible range. Other environmental factors were not considered for this kind of sensor. Regarding the biometric capture device location and considering that the sensor was a "desktop device" (i.e. a computer peripheral), it was placed in a standard table straight for all scenarios except for "scenario 2" that was tilted with an angle of 40°.

- Test crew. Test crew size was selected considering the supplier claim, the rule of three with a 95% of confidence level, and the time and effort that entails to find test subjects. According to the supplier, the sensor had a FRR = 0.34% (for FAR = 0.00%); hence, applying the Rule of 3, it means that, at least, 883 genuine comparisons have to be executed. It was decided to perform the same number of impostor comparisons. Due to the difficulty of finding this number of persons and considering that this evaluation aim was only the assessment of the proposed methodology, this quantity of comparisons was achieved but using samples of the same person (not totally independent). As a result, 10 individuals were recruited and each person provided his middle and index finger of both hands. Such biometric characteristics were chosen because these are the fingers recommended by the supplier. In order to satisfy the number of comparison obtained by the Rule of 3, each person had to perform 5 genuine and 5 impostor comparisons per visit and the number of visits was 2. In case of "scenario 5" test subjects conducted their transaction after applying hand cream to their whole hands.

  Moreover, test subjects were instructed by the test operator during the first visit for enrolment and verification processes. To explain how to use the device, user guides provided by suppliers were used and the test operator showed correct vs. incorrect usage with some examples. Also test subjects took part in some practical trials before carrying out the real transactions. For "scenarios 3 and 4", specific instructions were explained about how to present the biometric characteristic to the sensor with the corresponding movement up or down.

  Regarding feedback, visual feedback was shown to test subjects during their interactions with the device and at the end of each attempt. This way, they knew the image quality when they presented their biometric characteristic and also the match decision.
- Level of effort and decision policies. The level of effort and decision policies were the following:
  - Two maximum transactions for enrolment. Only if the first transaction fails, the second transaction must be executed.
  - Three attempts per each verification transaction.
  - The maximum time was limited to 60 seconds for enrolment and 5 seconds for verification.
- Error protocols. Just those errors involving a wrong identification or that the sensor did not work properly were considered.
- Data to record and test results. For that experiment only traditional performance metrics were obtained.

For conducting the complete evaluation, a specific application was developed for recording all data generated during the evaluation. This application was installed in a Core2 duo U7600 laptop with 1.2 GHz, 2GB of RAM whose operating system was Windows XP.

After the completion of the test plan and the end of test subject interactions in all "scenarios", biometric performance metrics were obtained. Results can be seen in Table 22 for enrolment and in Table 23 and Figure 18 for verification. In spite of this being the first version of the H-B interaction methodology, it was possible to analyse and quantify the influence of the assessed factors.

As it can be seen, the most influential factor for the tested biometric system was that users do not present their fingerprint in the centre of the scanner area. Other factors such as hand cream and a variation of the position of the biometric system device also reduced biometric system performance but such reduction is lower than the one dealing with the misplacement of the finger.

Table 22. Results obtained for enrolment [FER'10b]

Enrolment results		Common Scenario
FTE		5% first enrolment 0% second enrolment
Time to enrol	Average	16.282 s
	Minimum	8.15 s
	Maximum	82.86 s
	Standard Deviation	11.728 s

Table 23. Results obtained for verification process [FER'10b]

Verification results		Scenarios				
		1	2	3	4	5
No. of samples		2,080	2,046	1,923	1,944	1,920
No. genuine comparisons		1,056	1,026	978	982	960
No. impostor comparisons		1,024	1,020	945	962	960
FTA (%)		0.913	0.684	8.736	18.46	0.365
Time to capture (s)	Arithmetic mean	2.02 ± 0.52	2.15 ± 0.56	1.92 ± 0.55	1.84 ± 0.57	1.79 ± 0.39
	Minimum	1.14	1.19	0.68	0.93	0.91
	Maximum	8.48	7.65	9.41	10.43	4.53

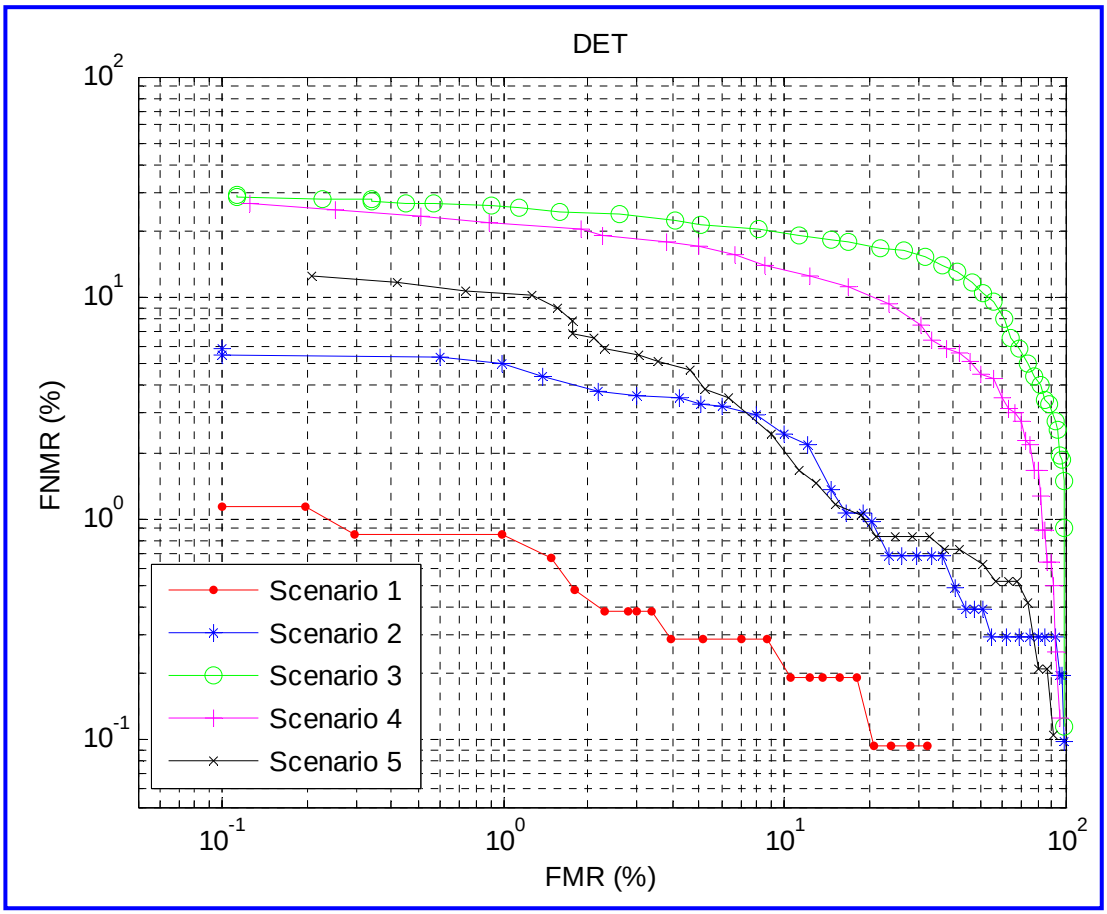

Figure 18. DET curve for all assessed evaluation conditions [FER'10b]

6.7.2 Development of the evaluation methodology and further experiments for improving it

Based on this first experiment, a second experiment was developed with the intention to improve the first version of the methodology. Specifically, it covered the analysis of several biometric systems at the same time and the recruitment of a more significant test crew which included test subjects of different ages and skills.

The purpose to carry out an analysis of several biometric systems at the same time was similar to the already mentioned for the development of the environmental testing methodology. This fact implied the definition of additional requirements for completing the evaluation methodology.

Moreover, the objective of recruiting a significant test crew covering different ranges of age and skills was to fulfil those requirements related to training and guidance of the test crew. For the H-B interaction methodology this aspect is fundamental due to its relation to test subject behaviour.

As a consequence, five fingerprint verification systems that have been developed by different companies were tested following similar test procedures. The test crew was composed of users of three groups of ages: 18 to 30, 31 to 50 and 51 to 65. In each group, the number of males and females were homogenous having distribution between 40% to 60% of men and women. Besides, the level of technical knowledge was different through such users.

Regarding the evaluation of multiple biometric systems, it was necessary to align developers' recommendations for defining the test plan and to design a proper test order considering factors such as to reduce the transaction duration, to avoid that subjects would get tired or confused, and to avoid habituation effects. Also specific guidelines and training were implemented in order to instruct test subjects for interacting with the different biometric systems.

In relation to managing test subjects of different ages and technical knowledge, it was essential to adapt the training to each test subject. It was also necessary to modify test procedures to reduce the duration of test subjects' visits. A mock evaluation was performed for checking all test procedures and it was detected that a visit may take too much time for non-habituated users.

Unfortunately, this experiment is the subject of a non-disclosure agreement and it is not possible to provide details about the tested biometric systems and results achieved. Nevertheless, a brief description of the experiment explaining just the test methodology and the inconveniences faced in this kind of evaluations can be seen in the work titled "Usability Evaluation of Fingerprint based Access Control Systems" [FER'10a]. Moreover, the corresponding requirements and improvements were added to the proposed methodology.

However, during the execution of this second experiment, the work on HBSI metrics (i.e. [ELL'10]) was published. As already mentioned these metrics expanded the description of FTA rate and were considered fundamental measurements to be included at the proposed methodology.

For this reason an exhaustive analysis of the complete HBSI model and the evaluation metrics proposed was carried out with the intention to review the second version of the methodology and include the HBSI metrics on it. This analysis was developed at Purdue University’s Biometric Standards, Performance & Assurance Laboratory under the support and supervision of Prof. S. Elliott who provided his thoughts and helped to clarify doubts.

As a consequence, a comprehensive revision of the methodology was done including the following advances:

- The concept of usability was changed to the broader concept of "H-B interaction" which covered not only usability but also ergonomics.
- A new factor classification was developed. This entails the current organization in three groups based on the components that compose the HBSI model. Also this current classification addresses usability and ergonomic factors.

- HBSI metrics and how to obtain them were added to the methodology. In addition other measurements have been described to analyse usability, ergonomics and signal processing aspects apart from the biometric system performance.

Therefore, multiple modifications were applied to the methodology and the current version, which has been explained in this chapter, was obtained. Nevertheless, further researches need to be performed in this area, essentially for completing this version with requirements and test procedures for analysing those factors that were not possible to be covered in this dissertation (see section 6.2.2).

6.8 Conclusions

This chapter has explained an evaluation methodology to analyse the influence of H-B interaction factors in biometric performance with a double intention. On one hand to fulfil other factors that belong to the environment that were not covered during the specification of the environmental testing methodology (explained in the previous chapter). On the other hand, to provide an evaluation methodology for studying the influence of human factors on biometric system performance. In a similar way to ambient conditions, human factors are relevant parameters that traditionally have been claimed as having a high level of influence. However, no evaluation methodology was established at the time this work started.

In the same way as with the environmental testing evaluation methodology, the evaluation methodology described in this chapter has provided requirements for planning, conducting and reporting this kind of evaluation based on ISO/IEC 19795 multipart standard. Also, the philosophy for the evaluation model has been based on the previous published works on the HBSI model, which has been used as the starting point for developing the H-B interaction evaluation methodology. In a more precise way, this second evaluation methodology includes the specification of the following aspects:

- H-B interaction factors that may be analysed and how these conditions shall be specified for their evaluation. A detailed list of all possible human-biometric interaction factors have been defined, although it has been not possible to cover all of them.
- Specific requirements for carrying out H-B interaction testing of biometric systems considering a biometric performance scenario evaluation. Additional requirements about the environment, test crew, test procedures, the sequence of execution of the trials, error protocols, data to record and test results have been defined.
- The establishment of a baseline performance in order to accurately obtain the influence on biometric performance of the tested H-B interaction factors.
- Specific measurements and metrics for this kind of evaluations beyond performance metrics.

The specification of this evaluation methodology together with the methodology specified in Chapter 5 satisfies the objective of this Thesis of developing formal evaluation

methodologies for testing biometric systems working under specific contour conditions. Although both evaluation methodologies can be improved adding the specification of further ambient conditions to tests and/or completing the evaluation requirements for the entire list of H-B interaction factors, the basis for such research activities has been established.

The work in this field has provided the following set of publications:

- B. Fernandez-Saavedra, R. Alonso-Moreno, J. Uriarte-Antonio and R. Sanchez-Reillo, *Evaluation methodology for analyzing usability factors in biometrics,* Aerospace and Electronic Systems Magazine, IEEE, 2010c, 25(8), p. 20-31, 2010 [FER'10b].
- B. Fernandez-Saavedra, R. Alonso-moreno, A. Mendaza-Ormaza and R. Sanchez-Reillo, *Usability Evaluation of Fingerprint Based Access Control Systems,* 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 [FER'10a].

Chapter 7

Guidelines for conducting biometric performance testing according to CC and CEM

As it was mentioned in Chapter 4, Common Criteria is currently the only international recognised evaluation framework that biometric system developers can follow to analyse and demonstrate the level of security achieved by their products. However, the applicability of this methodology to biometric technology needs the specification of supplementary guidelines.

This chapter provides such guidelines with the objective that biometric systems can be accurately tested applying CC and CEM. First of all, an overview of the current situation and the necessity of the proposed guidelines are described. Then, a general biometric system is explained in the context of CC. After that, the CC testing activities that shall be carried out for analysing a biometric system are summarized noting the importance of conducting a biometric performance evaluation. As a result, the most relevant parts of CC related to biometric performance evaluations will be detailed indicating for which CEM evaluation activities are essential to specify additional guidelines. Once these activities are known, their corresponding work units, together with the proposed guidelines regarding biometric performance testing are given. It is important to highlight that these guidelines have been defined in compliance to the existing ISO/IEC 19795 standard and considering previous works developed in this area. In addition, an interpretation of contour conditions and their influence on biometrics is described from a CC point of view. Finally, the research works performed for the development of these guidelines will be explained.

7.1 Overview

In a short period of time, biometric systems have become indispensable in scenarios where human recognition and security are two critical requirements such as border control or banking. However, in spite of this kind of products being used more and more, and the significance of their proper working, their evaluation is not a common practice. This is due to the fact that evaluating biometric systems is a real challenge as it was described in Chapter 3. On one hand, the evaluation of biometric systems is a complex, costly and time consuming process. On the other hand, standard evaluation methodologies have been approved recently but a certification scheme does not exist yet. Nevertheless, due to the use of biometric systems, it is essential to demonstrate that these systems achieve an acceptable level of accuracy and security.

Currently there are two formal ways for testing biometrics products. One option is that biometric systems may be tested in accordance to international standards such as the multipart standard ISO/IEC 19795 which address biometric performance testing, or the ISO/IEC 19794 series which include the specification of data format conformance. However, this type of evaluation only covers the analysis of those requirements that are included on the relevant standard. Alternatively, Common Criteria is a certification scheme for assessing the security of IT products in which biometric systems are included. Regarding biometric technology, the analysis of security involves the analysis of the system capabilities for people identification. Therefore, CC encompasses the examination of the most relevant biometric system properties.

Considering these circumstances, CC entails a more exhaustive evaluation. Its evaluation approach entails an overall evaluation process that analyses security considering each of the different aspects related to the product (i.e. its design, development, delivery, documentation, operation and vulnerabilities). Nevertheless, as it was mentioned in Chapter 4, CC is general and it is not totally adapted to biometrics. In a similar way to other technologies, specific guidelines must be developed to help developers and evaluators to understand and apply them. Some works have already been done, but either these works need to be updated to the current version of CC and to biometric system performance evaluation methodologies like BTSE and BEM, or it has to be more thoroughly defined like in the case of ISO/IEC 19792.

In view of this situation, this dissertation proposes new guidelines for applying CC to biometric products. Although these new guidelines are based on the above mentioned works (i.e. BTSE, BEM and ISO/IEC 19792), these have been updated to the current versions of CC and CEM, and also considering current versions of ISO/IEC 19795 Parts 1 and 2. Nevertheless, these guidelines only cover biometric performance testing and the assessment of those threats that can be counteracted by achieving high levels of technical performance, such as impersonation and disguise attacks. A detailed methodology for analysing all kind of biometric vulnerabilities has not already been specified so it is not possible to propose a formal testing methodology for them. Furthermore, biometric system vulnerabilities that are common to other IT products can be assessed using the current CC evaluation methodology without any additional specification.

Exactly, these guidelines describe in detail specific requirements per each work unit of the CC evaluation activities involved in biometric performance testing, considering the following requirements:

- are independent of any biometric modality;
- are specified covering those kinds of biometric performance evaluations that can be repeatable, i.e. technology and scenario biometric performance evaluations;
- are based on previous works: BEM, BTSE and ISO/IEC 19792; and
- consider the last version of both CEM and ISO/IEC 19795 Part 1 and 2 standards.

Additionally, some considerations about how the environmental conditions and H-B interaction influence on biometric system performance shall be interpreted in terms of a CC security evaluation are provided.

7.2 Biometric systems as a TOE

For the purpose of these guidelines and in terms of CC, a biometric system will be the TOE. This TOE has the capability of recognising people by means of physical or behavioural characteristics that they posses. Considering this definition both, the security problem definition for this type of TOE and the related security objectives to solve this problem, are described as follows.

7.2.1 Security problem definition

For this type of TOE, the primary threats to these systems are those of an impostor gaining access to the assets under protection, as well as the fact that an authorized user is not being recognized and, therefore, cannot access to such assets. These systems have further threats which affect to the information and resources that these systems use during their operation such as the biometric reference, personal data to the user, quality and decision thresholds, recognition algorithm, etc. Attacks that modify or steal them or attacks that inject a new one are also threats which should be part of the security problem definition of a biometric system. However, these threats are common to other IT products, so for the intention of these guidelines these are not going to be included. Furthermore, there are other threats that affects to users, as they can be threaten to present their biometric characteristic. Nevertheless, these are also common to other authentication technologies and are not to be included either. A complete list of all potential threats for a biometric system is detailed in [BEM'02].

7.2.2 Security objectives and its implementation

Consequently, regarding the primary threats, the security objectives to achieve by biometric systems will be authenticating/identifying individuals correctly and fulfilling specific error rates. The latter objective is due to the probabilistic nature of biometrics. These objectives could be described using different security functional requirements (SFRs), but they

are usually implemented by means of two general functions: enrolment and verification/identification. Generally, these functions include the following activities:

- Enrolment function. This function collects and stores the biometric reference from the subject. For doing that, the administrator has to execute this function and provide user's information. Then, the subject must present his biometric characteristic. If everything is correct, the result will be that the biometric template of such subject has been correctly generated and stored. Otherwise, the result will be a failure to enrol.
- Verification/Identification function. These functions are responsible of recognizing individuals correctly. First, the individual has to execute the function. In the case of verification, he also must claim his identity. Then, he/she shall present his/her biometric characteristic to the capture sensor. Finally, the biometric system gives back the result of the recognition process. For a verification function this result is a comparison score or an accept/reject decision whereas for an identification function this result is a candidate list.

Furthermore, most of the biometric systems have the possibility to configure certain parameters of these functions such as quality thresholds, decision thresholds, maximum number of attempts, time out for capturing the biometric characteristic, etc [LI'09]. Therefore, a function that allows the configuration of those parameters shall be considered as a biometric function.

Considering these security objectives and their implementation, the TOE Security Functionality (TSF) of a biometric system is the combination of the hardware and the software that are involved in the enrolment and verification/identification functions. Typically, the hardware is composed of a biometric capture device and the necessary circuitry upon which the software operates, including the storage media used within the process. On the other hand, software consists of the collection of programs which implement algorithms and its information to perform such functions.

During a CC evaluation, developers will have to specify the design and implementation of the SFRs at various levels of abstraction (functional specification, TOE design and implementation representation). This helps evaluators to get a better understanding of the TOE for the subsequent testing activities. Regarding biometric systems, there are a lot of possibilities to design and implement this type of TOE. Several biometric modalities exist and most of them could use different kind of algorithms. Moreover, the representation is reliant on which parts of the biometric system are going to be included as part of the TSF and its complexity.

Nevertheless as a reference for the specification of the current guidelines, it is indispensable to define these aspects. Therefore, a common representation of a biometric system has been described for the highest level of abstraction including the TOE design and the functional specification. This representation is based on the general model of the biometric

system that was established by ISO/IEC JTC1 SC37 in its Standing Document SD11 [ISO'10a] which was described in Chapter 2. This model has been slightly modified to be consistent with CC resulting in the model shown in Figure 19. The modification consists of taking the presentation of the biometric characteristic out of the biometric capture subsystem in order to make it possible to delimit the TOE boundaries.

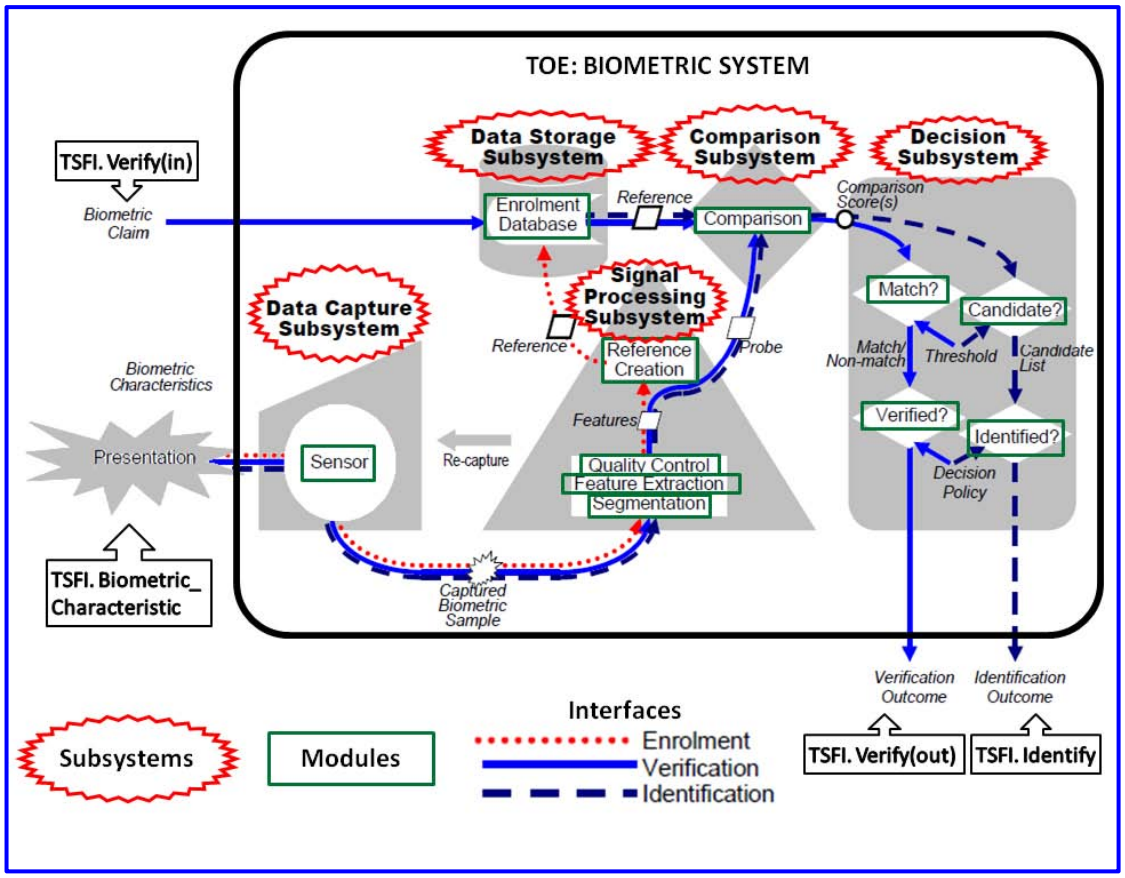

Figure 19. TOE Design of general biometric system [ISO'10a]

On one hand, the TOE design entails the description of the TSF considering two levels of decomposition: the subsystems and the modules in addition to their communications (i.e. the interfaces). Figure 19 shows those elements in case of a typical biometric system: data capture, processing, data storage, comparison and decision. These subsystems have been highlighted in bold inside an oval. In addition, the figure shows the possible modules that make up each subsystem. These modules have been indicated with a rectangle. When the functionality of these modules will be very complex, a decomposition using further subsystems levels may be needed. Then, the lower-level subsystems shall be divided into modules. Interfaces between modules are depicted by means of arrows of different types. The features of the arrows represent the purpose of their interaction, which is performing the enrolment, verification or identification functions. It is important to emphasize that Figure 19 does not include the configuration subsystems and their corresponding modules but it should be also described if TOE has this functionality.

On the other hand, the functional specification of the TOE describes the TSF interfaces (TSFIs) explaining the way that external entities receive data from the TSF, supply data to the TSF or the TSF services that are invoked. As it was mentioned in previous paragraphs, the TSF of a biometric system covers three main functions: enrolment, verification/identification and configuration. Regarding these functions, the TSFIs of a generic biometric system are made up by two groups: physical and logical. Next, each of these interfaces is going to be described in a generic way, in accordance to CC. Besides, some of them have been depicted in Figure 19.

- Physical interfaces: Biometric sensor.
  - TSFI.Biometric_Characteristic. The purpose of this interface is to obtain a signal that represents the biometric characteristic of the biometric capture subject. This subject must interact with the biometric sensor and present his biometric characteristic. The parameters of this interface depend on the type of sensor. Most of them provide feedback by means of visual and/or audio indications which guide subjects in their interactions. The action of this interface is to acquire a suitable sample of the biometric characteristic. Error messages will be messages indicating a failure during the capture process that has been performed, i.e. the "Failure to Capture" error.
- Logical interfaces: Biometric Functions.
  - TSFI.Enrol. This interface invokes the enrolment function. The administrator initiates this process, for example by clicking the "Enrol" button of a Graphical User Interface (GUI). The input parameters are a personal identifier and personal data (e.g. name, surname, age, sex, etc.), whereas the output parameter is the result of the enrolment process. Its action is to enrol individuals in the biometric system. The error messages are messages that indicate a "Failure To Enrol" (FTE) error.
  - TSFI.Verify. This interface invokes the verification function. It is initiated by the user, for example by clicking the "Verify" button of a GUI, introducing a PIN, etc. The input parameter of this interface is a personal identifier and the output parameter is the result of the verification process. Its action is to check the claimed identity of an individual. The error messages are messages which report a "Failure To Acquire" (FTA) error when the biometric sample cannot be acquired or a "failure to compare" error, when the comparison process cannot be completed by any reason.
  - TSFI.Identify: The purpose of this interface is to invoke the identification function. The user initiates this process, for example by clicking the "Identify" button of a GUI or interacting with the biometric capture sensor. In contrast to TSFI_Verify, this interface does not need to provide a personal identifier, so there are no input parameters. The output parameter is the result of the identification process. The error messages are messages which report a "Failure To Acquire" (FTA) error when the

biometric sample cannot be acquired or a "failure to identify" error when the identification process cannot be completed for any reason.

- TSFI.Configure: This interface invokes the configuration function. This process is initiated by the administrator, for example by clicking the "Configuration" option of a GUI. Its input parameters are parameters such as the number of attempts per transaction, timeouts, quality and decision thresholds, etc. The error messages are messages which indicate that the value is not possible or the value could not be set.

7.3 CC testing activities for biometric systems

In a CC evaluation, the testing activities consist of analysing that the TSF behaves accordance with its specification and checking that the TOE does not have exploitable vulnerabilities.

Regarding biometric systems there are a lot of aspects that compose a security evaluation. However, most of them can be analysed following similar approaches than other IT products such as its development, implementation and the analysis of certain vulnerabilities (e.g. malware, data injection, communications interception, hill climbing attacks, etc). Only aspects associated to biometric technology functionality and their particular vulnerabilities need an extra explanation.

In relation to the analysis of vulnerabilities, a formal methodology to analyse them has not been specified yet. There are some works that have started to be developed: either by ISO/IEC JTC1 SC37 [ISO'12c], CC [CCN'08] and other institutions [SAN'06, FER'08b, HEN'10, MUN'12]. Nevertheless, this methodology shall be specified and adapted to each biometric modality and the particular technology of the biometric capture device used for the implementation of such modality.

As a result, the proposed guidelines are focused on the calculation of biometric system technical performance rates and on those requirements involved in it. The main reason is because measuring this aspect can not only demonstrate a correct behaviour of the most significant security property of a biometric system, but biometric performance also quantifies the probability that biometric system has to face impersonation and/or disguise threats. Examples of such impersonation threats are "zero effort" impostor attempts attacks, blood relationship impersonation attacks or "wolf" biometric sample attacks [DOD'98], whereas examples of such disguise threats are when people modify their biometric characteristic in order to avoid being recognized, also known as obfuscation. In addition, a preliminary measurement of biometric performance is essential as a reference to quantify the effects of other potential threats such as environmental conditions or modifications of biometric characteristics among others. This decision is also based on those previous published works explained in Chapter 4 (i.e. BTSE [BTSE'01], BEM [BEM'02] and ISO/IEC19792 [ISO'09a]). All these documents present the common idea that a biometric system security evaluation shall include the calculation of the most relevant performance rates, which are so-called error rates,

i.e. the False Non-Match Rate (FNMR) and the False Match Rate (FMR). These documents define that error rates are significant measurements of the security of biometric technology. According to BTSE, FMR quantifies the "zero effort" impostor attempt vulnerability, whereas FNMR quantifies the availability provided by biometric systems. In the same way, ISO/IEC 19792 defines FMR as a security value and FNMR as a usability value. Furthermore, all these works remark the relationship between these metrics and the importance to analyse and report them together. Moreover, most of the existing dependent and independent statements of security needs for biometric devices in CC (i.e. one Security Target [ST'08] and two Protection Profiles [PP'05, PP'08]) point out the importance of biometric systems meeting specific performance rates. These documents define an organisational security policy which determines values for such error rates. Then, this policy is translated into a refinement of the functional requirement FIA_UAU (i.e. Functional requirement of Identification and Authentication class which addresses User AUthentication) in order to guarantee that the achievement of those error rates values is checked.

Considering the calculation of performance metrics, the existing methodology that establishes how to obtain them is the international multipart standard ISO/IEC 19795. As it was explained in Chapter 3, the Part1 of this standard classifies the performance evaluation of biometric systems in three types: technology, scenario and operational evaluations. However, not all types of performance evaluations are consistent with CC. CC claims objective and repeatable results, so only technology and scenario evaluations are going to be considered at the proposed guidelines. For operational evaluations it is not possible to control most of the parameters. Strictly speaking, scenario evaluations are claimed to be repeatable only to the extent that the modelled scenario (i.e. environment conditions and human factors variables) can be carefully controlled. However, due to this type of evaluations covering the evaluation of the complete biometric system, these are going to be included addressing also the necessary requirements for an exhaustive control the influential factors.

In the following sections, this methodology is presented according to CC. First, a description of the assurance classes involved in the biometric system performance evaluation is provided. Then, the testing guidelines are specified per each assurance class according to CEM requirements as well as the TOE design and TSFI specification described here.

7.4 Security Assurance Requirements and ISO/IEC 19795

As it has been explained in Chapter 4, CC is a standard composed of a set of functional and assurance requirements. Functional requirements are essential conditions that determine the security functionality of the TOE. Depending on the product under evaluation and the security objectives specified for such product, the security functionality of the TOE shall meet some of these functional requirements. On the other hand, assurance requirements are assurance measures to be applied during the security evaluation. Depending on the evaluation assurance level (EAL) chosen, a subset of these assurance requirements shall be analyzed by evaluators.

Both functional and assurance requirements are hierarchically organized from the higher to lower level into classes, families and components.

Assurance classes are defined considering all aspects that have to be analyzed in a CC security evaluation. The PP evaluation class (APE) and ST evaluation class (ASE) assess the statement of security needs which are claimed for the type of TOE or a specific TOE respectively. Both its contents and consistency are analysed. Development class (ADV) checks that the security functionality of the TOE works and cannot be corrupted or bypassed. For doing this, this class analyses the functional specification as well as the design, implementation, interfaces, architecture and internal structure of the TOE. Besides that, Guidance Documents class (AGD) examines the documentation to handle the TOE in a secure way during its preparation and operation. Life-cycle Support class (ALC) establishes the requirements for controlling the process during the development and maintenance of the TOE. Moreover, Tests (ATE) class addresses functional testing of the TOE to check that it behaves correctly. It consists of analyzing the operation of the TSF and the TSF interfaces (TSFIs). In addition, Vulnerability Assessment (AVA) class addresses the analysis that no exploitable vulnerabilities exist. Finally, there is a Composition class (ACO) for testing composed TOEs.

Nevertheless, not all assurance classes are involved in a biometric performance evaluation. According to ISO/IEC 19795-Part 1, biometric performance evaluation requires planning and execution of several activities. In order to plan these evaluations it is necessary to know information about the biometric system and the objectives of the evaluation. Test execution entails pre-test activities including installing the biometric system in the environment, checking its correct operation, recruiting test subjects (or preparing a database), training test operators and implementing the evaluation application for recording test data automatically. Then, during the execution of the tests, enrolment and verification processes have to be carried out. Also, in some cases, test subjects must be trained to use the biometric capture sensor before those processes take place. At the end, evaluators shall obtain performance metrics and uninstall the biometric system, only keeping the relevant data to keep traceability and repeatability of the test carried out.

Consequently, the assurance classes concerned with the calculation of biometric performance are Guidance Documents and Tests (AGD and ATE). Requirements related to the installation and operation of the biometric system, as well as the training of evaluators and test subjects shall be added to the AGD class. Likewise, ATE class must include further requirements related to the way of planning, executing and reporting biometric trials and results.

Both classes will be covered by the proposed guidelines as it is shown in Figure 20. Such guidelines will be described in detail in the next sections considering the highest assurance components of CEM. It is important to note that these activities will be the previous steps to apply AVA class, since this class entails the assessment of potential vulnerabilities that can affect biometric performance but not the establishment of biometric performance itself.

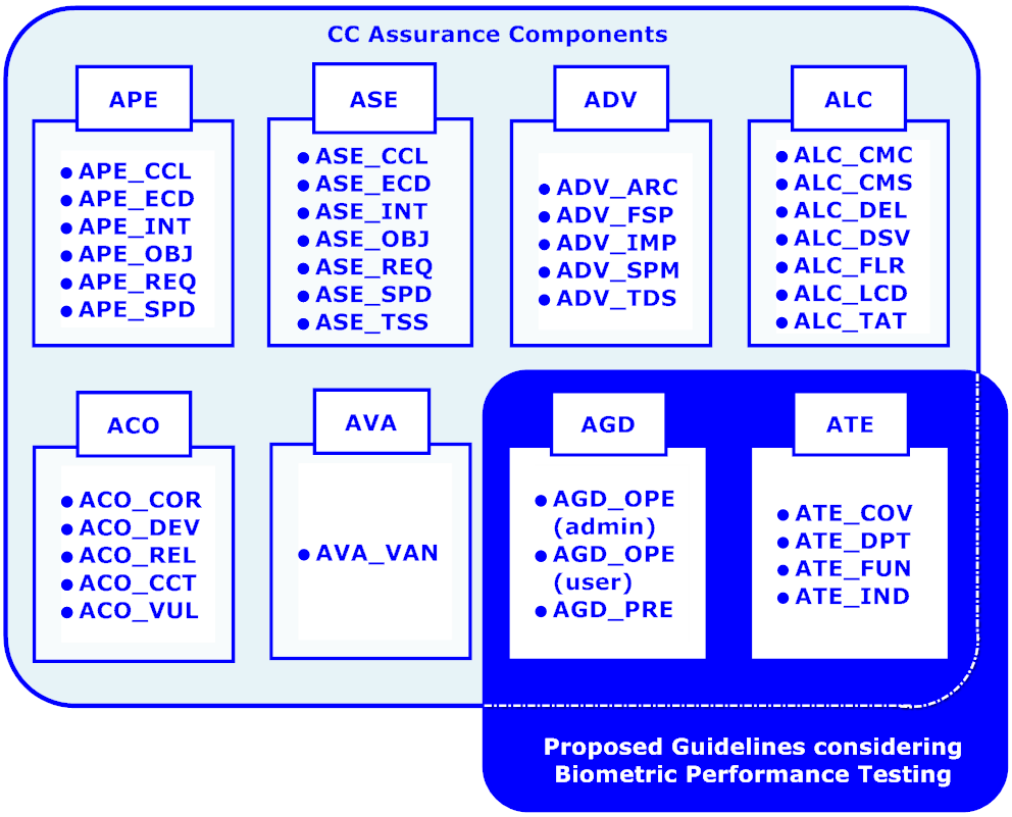

Figure 20. Assurance components covered by the proposed guidelines

7.5 AGD Class: Guidance documents

AGD class addresses the analysis of the installation and operation documentation by evaluators. For a biometric TOE, guides must include information about critical aspects such as the proper environment and the correct interaction with the biometric capture sensor.

AGD has suffered many changes from version 2.1, for which BEM and BTSE were developed, to the current version of CC and CEM. Table 24 summarizes these changes. In the new version, the previous Configuration Management (ACM), Delivery and Operation (ADO), Life-Cycle Support (ALC) and Guidance Documents (AGD) classes were mapped just in two classes: ALC and AGD. Therefore, the new version of AGD class includes components of the old ADO and AGD classes.

Table 24. Differences between versions of CC

Version 2.1	Version 3.1 Revision 4
ADO_IGS: Installation, generation and start-up	AGD_PRE: Preparative procedures
AGD_ADM: Administrator guidance	AGD_OPE: Operational user guidance
AGD_USR: User guidance	

As a result, the new guidelines have to be modified considerably, although certain guidelines provided by BEM and BTSE are still applicable. Besides, some requirements provided by ISO/IEC 19792 also apply. This has been illustrated in Figure 21. This figure depicts the current AGD families and the guidelines that correspond to each family. These guidelines have been classified depending on its sources: the previous works and the guidelines proposed at the current research work. Next, the specific considerations regarding families that belong to AGD class will be described.

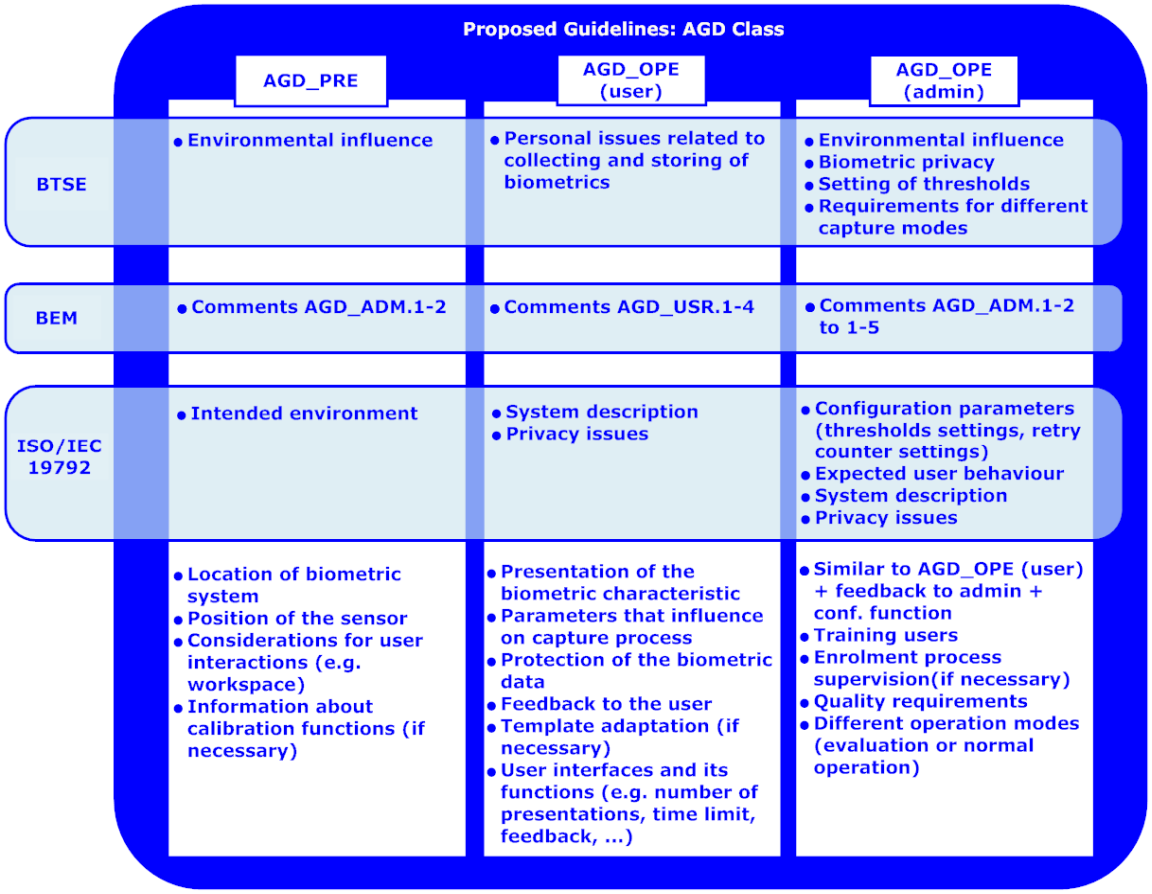

Figure 21. Proposed guidelines for AGD class

7.5.1 AGD_PRE: Preparative procedures

This family addresses how to evaluate the description of preparative and delivery procedures provided by developers. Regarding biometric performance testing, the most significant activity is the preparation of the TOE in its operational environment (AGD_PRE.1-2). As the operational environment could affect the process of acquiring the biometric characteristic, some installation guides are essential to reduce the effect of the environmental conditions on the TSFI.Biometric_Characteristic. BTSE and BEM documents establish that guidance documentation should include information about the influence of environmental factors and the ways to minimize it. This is also a requirement of ISO/IEC 19792. Nevertheless, the operational environment for a biometric system should not only cover environmental factors, but also the location of the system and how individuals shall interact with the biometric device.

Regarding the biometric capture sensor position, developers shall provide advice related to height and orientation. If multiple locations and deployments are considered (e.g. table, wall, counter, turnstile, etc.) this documentation shall provide information about all of them. Likewise, developers must describe considerations for interactions. If a biometric system needs particular workspace or has to be placed in a specific side, such specifications have to be provided. If an individual has to present his right hand and the biometric system is placed in a turnstile in the left side, it is likely that the biometric capture subject turns his hand or himself and does not provide a good sample. Factors like that reduce biometric performance.

Furthermore, there are biometric systems that have specific functions to adjust their operation to the existing environmental factors (e.g. optical sensors have a function to calibrate the illumination level). In this case, preparative guidance should include the explanation of these functions, specifying the way and time to execute them. In the same way, if these functions are considered part of the TSF (i.e. TSFI.Configure) and/or should be executed during the TOE operation, this information shall be included in the administrator operation guides. It will be described in the next component.

7.5.2 AGD_OPE: Operational user guidance

Operational user guidance family is applied to assess that the related documentation describes the functionality and interfaces of the TOE, as well as how to use it in a secure way for each user role. In its previous version, this family had two components, one for user guidance and other for administrator guidance, so BEM and BTSE specified guidelines for both of them. Now, this division does not exist anymore, but the current version requires specifying the operational guidance for each user role. A biometric system has typically two authorized kind of users: biometric data subjects (which are similar to the group called "users" within previous works) and administrator. For this reason, the same division of previous works will be kept. Next, it will be detailed the information to incorporate for completing the information given for previous works considering the biometric TOE and the proposed TSFIs in section 7.2.2.

7.5.2.1 Biometric data subject guides

Regarding this issue, BEM document is more specific than BTSE. BEM states that user guidance should include guidance on the capture and enrolment processes and on aspects related to personal issues such as privacy. All these statements are correct, but this description should be more detailed. Besides, it is necessary to add other guidelines that were not mentioned in BEM. These guides will be the following:

- Physical Interface: TSFI.Biometric_Characteristic. The guidance for the capture process must describe how to present the biometric characteristic to the capture sensor (AGD_OPE.1-2). It shall also inform about any physical element, behavioural aspect or chemical product that could influence on such process, such as removing glasses, avoiding laughing or trying to avoid the use of hand cream or make up (AGD_OPE.1-2, AGD_OPE.1-3 and AGD_OPE.1-6). All these factors will depend on

the biometric modality of the TOE. ISO/IEC TR 19795-3 [ISO'07c] lists them for each modality. Moreover, the biometric data subject operational guidelines have to offer recommendations about the way to protect biometric data (e.g. wiping off fingerprints of the capture sensor area) or other parameters or sensible data of the biometric system (e.g. not tampering with the device, or how to activate an alarm in case of coercion or other emergencies).

Other aspect to be included in the biometric data subject guides regarding this interface is the feedback these people receive when they interact with the biometric capture sensor (AGD_OPE.1-2 and AGD_OPE.1-4). This feedback could be given during and/or after the process. Biometric data subjects must know what happens during their interactions and the actions to perform at any specific event.

Furthermore, there are biometric systems that adapt the template after a successful verification. This entails a modification of the security parameters. As biometric data subjects should be aware of the change, operational guides shall describe this process (AGD_OPE.1-4).

- Logical Interfaces: Biometric Functions. Biometric data subject guidelines have to explain each of these interfaces (GUIs) and their corresponding functions (AGD_OPE.1-1 and AGD_OPE.1-2). This description must include the number of times to present the biometric characteristic, if there is a limit of time, the provided feedback, as well as how biometric data subjects should act in any case (AGD_OPE.1-3 and AGD_OPE.1-4).

  About the enrolment process (TSFI.Enrol), biometric data subject guidance also has to explain which personal data should be provided by individuals and how these data and biometric data will be handled to guarantee privacy. For specific applications with the assumption of unique enrolment, biometric applicants must prove their identity before the enrolment process. They must know this requirement (AGD_OPE.1-6). Likewise, regarding the verification process (TSFI.Verify), this document shall describe the way that biometric capture subjects have to provide their identifier to this function (AGD_OPE.1-2).

7.5.2.2 Administrator guides.

Regarding these guides, BTSE and BEM documents require the description of environmental controls and how environmental factors affect the security of the system. Such documents also address the consideration of decision thresholds as security parameters and state that administrator guidance should consider user behaviour and the need for biometric data subjects to be monitored or supervised. In addition to these requirements, the recommendations to be added to the administrator operational guide are the following:

- Physical Interface: TSFI.Biometric_Characteristic. The administrator guides have to be similar to the user guides mentioned above. Also, these guides must explain the procedures related to the way administrators have to train biometric data subjects for the biometric interactions and the quality requirements to determine correct

or incorrect presentations of the biometric characteristic (AGD_OPE.1-2 and AGD_OPE.1-3).

- Logical Interfaces: Biometric Functions. These functions and their interfaces must be explained in the same way as biometric data subject guides. The enrolment process also has to cover quality requirements to accept or reject an acquired sample before the template creation, when this decision is not automatically made. In addition, this guidance shall specify how administrators have to deal with both personal and biometric data of the individuals. On the other hand, in the verification/identification process, it must be described whether administrators shall supervise these processes.

Additionally, the administrator operational guide has to include the explanation of the configuration function and its interface (TSFI.Configure). In particular, this guidance must specify how to manage security settings such as quality thresholds, decision thresholds, maximum number of attempts, maximum number of transactions, maximum time for attempt or transaction, maximum number of identifiers to include in the candidate list, etc. (AGD_OPE.1-3). Also, any environmental suggestion has to be described as it mentioned in Preparative Procedures (AGD_OPE.1-6). It is important to highlight that enrolment and verification/identification processes may require different configurations. The specific settings for each process shall be explained in this guide.

Furthermore, there are biometric systems that have a variety of operation modes. This is very common in this type of systems because some essential information for performance evaluations (i.e. scores, times, additional user information) could be a potential vulnerability during the operation. In this case, administrator guides must specify the different operation modes (AGD_OPE.1-5).

Finally, it is important to emphasize that all information provided in the guidance documentation will be useful not only for training evaluators during the examination of this documents, but also for planning the biometric performance testing.

7.6 ATE Class: Tests

ATE class establishes the procedures that developers and evaluators have to carry out in order to check that the behaviour of the TSF meets the SFRs. For a biometric TOE, the analysis of the TSF entails to verify the correct operation of the enrolment and verification/identification functions as well as the fulfilment of specific error rates. Due to the structure of this type of TOE, such analysis requires the evaluation of several subsystems, as well as the corresponding modules and interfaces. According to CEM methodology, the more suitable procedure for testing the TSF of this type of TOE is to stimulate the TSFIs and observe their responses. So, biometric systems TSFIs must be stimulated following biometric performance testing methodology. From the results, it will be possible to calculate biometric system performance and the most significant error rates.

In order to complete its purpose, ATE is composed of four families, similar to previous versions. Two families specify requirements for the coverage (ATE_COV) and depth (ATE_DPT) of tests, other family addresses the functional tests carried out by developers (ATE_FUN) and the last family covers the independent tests performed by evaluators (ATE_IND).

The guidelines offered by BEM and BTSE documents for this class are quite similar and only cover Functional Test and Independent Test families in a general manner. Both of them address how to carry out performance testing and obtain FMR and FNMR error rates. Such rates shall be calculated using appropriate and statistically representative data and taking care of the collection procedure. The equipment and environment shall be configured properly and its functioning shall be verified previously to data processing. These recommendations are also requirements of ISO/IEC 19792. Figure 22 shows a diagram for the new guidelines including those requirements and the new ones proposed for this class. The ATE_DPT family is essentially empty because previous works have not provided any guidelines for it. Next paragraphs describe all of them in detail, according to ISO/IEC 19795 Parts 1 and 2 and considering all ATE families and their specific work units.

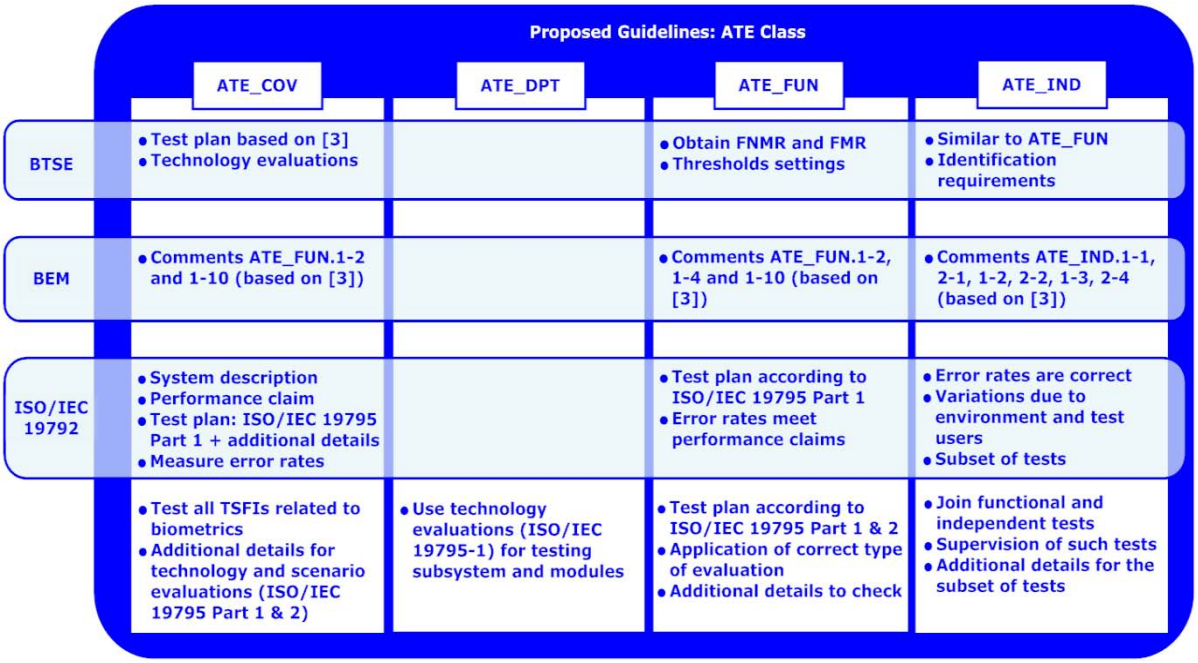

Figure 22. Proposed guidelines for ATE class

7.6.1 ATE_COV: Coverage

This family determines the TSFIs to be tested by developers. Such TSFIs shall be the same that were described at the functional specification. Therefore, the TSF interfaces to check for the proposed biometric TOE have to be the TSFIs specified in section 7.2.2 (ATE_COV.2-1).

Furthermore, this family addresses that evaluators shall examine the test plan to determine if the testing approach is suitable to analyse the behaviour of each interface (ATE_COV.2-2). As mentioned in section 7.3, ISO/IEC 19795 provides the best methodology to test the security property of correctly identifying individuals. However, it entails to check

several TSFIs and parts of the TSF at the same time. Considering either ISO/IEC standard or CC, two types of performance evaluation can be applied: technology or scenario evaluations. A technology evaluation shall be used when the biometric system under evaluation does not include the biometric capture sensor. Otherwise, developers and evaluators shall follow a scenario evaluation being strict with the control of environment and human factors variables. Apart from the test plan, evaluators shall examine the test procedures (ATE_COV.2-3). It means that evaluators have to check the test pre-requisites, test steps and expected results. For a biometric performance evaluation, the most important requirements are:

- Database size (technology evaluations) or test crew size, number of samples, number of transactions and number of visits (scenario evaluations), shall be chosen according to the expected error rates and confidence level to achieve. ISO/IEC 19795-1 establishes specific rules for it (rule of 3 and rule of 30).
- User characteristics must be consistent with a standardized corpus (technology evaluations) or the target population (scenario evaluations). For scenario evaluations, the composition of the test crew in relation to age, gender, race and skills shall be carefully specified because these factors may affect results.
- Environment has to be the target operational environment specified in the ST and shall be established in accordance with Preparative Procedures guides.
- Genuine and impostor transactions shall be performed to test TSFI.Verify in order to calculate the mandatory error rates. In case of TSFI. Identify, the type of transactions depends on the type of identification. If it is an open-set identification (i.e. all kind of people may utilize the biometric system), both genuine and impostor transactions shall be carried out. However, if it is a closed-set identification (i.e. only biometric data subjects may use the biometric system) only genuine transactions are needed.
- Test order shall be as follows:
    - TSFI.Configure. This interface must be tested firstly because it sets significant parameters for biometric performance testing such as maximum number of attempts, maximum time per transaction, quality and decision thresholds and number of attempts per transaction.
    - TSFI.Enrol has to be tested before TSFI.Verify or TSFI.Identify to generate templates prior to perform comparisons. ISO/IEC 19795-1 requires that enrolment is separated from verification and identification as much as possible for scenario evaluations.
    - Genuine transactions shall be carried out before impostor transactions for systems that adapt the template after a successful verification.
    - Impostor transactions to test TSFI. Identify based on an open-set identification function shall be conducted by test subjects that have not been previously enrolled. It may require having a dedicated test crew for this type of transactions.

Regarding test order, it is essential that developers check that the evaluation software saves the indispensable data to obtain error rates before test subjects perform biometric interactions. Otherwise, the data collection is useless.

There are further requirements that must be considered in this kind of evaluations according to ISO/IEC 19795 Part 1 and 2. Such requirements are those related to data preparation and corpus validation in case of technology evaluations as well as procedures for instructing test subjects, training them, habituation and acclimatization in case of scenario evaluations. In addition, it must be considered requirements for recording and reporting information in both cases. Developers and evaluators have to review such standards and apply them taking into account the specific TOE, its application and the criteria defined in the Operational User Guidance.

7.6.2 ATE_DPT: Depth

This family determines how developers shall test subsystems and modules that compose the TSF and how evaluators shall examine the test documentation to assess such tests. This activity is highly dependent on the TOE design and its architecture. However, ISO/IEC 19795-2 recommends technology evaluations to analyse all the subsequent parts of a biometric system after the human-sensor interface block.

7.6.3 ATE_FUN: Functional tests

Functional test family addresses the way developers shall conduct and report tests, as well as the way evaluators have to analyse if such tests have being performed and documented appropriately. When evaluations deal with biometric performance testing, the work units of this component shall be interpreted as follows.

Evaluators must check that test documentation includes the mandatory requirements addressed in ISO/IEC 19795 Part 1 and Part 2 for planning, executing and reporting biometric performance evaluations (ATE_FUN.1-1).

Evaluators shall identify the type of performance evaluation (ATE_FUN.1-2) and the approach to carry out each test.

- Technology evaluations. These evaluations are suitable for TOEs which do not include the capture sensor. In this case, the scenario description means specifying the approach by means of the biometric system processing samples from a database. Usually, this kind of evaluations needs to implement additional software for executing tasks such as selecting samples of the database as templates or samples, sending these samples to the biometric algorithm, recording times of processing and comparison, recording comparison scores and calculating performance rates. Also, it is essential that developers describe the database used in the evaluation. This database could be public or specifically created for the

testing purpose. In the latter case, the scenario specification must include data collection procedures.

- Scenario evaluations. These are the evaluations to be performed when the TOE includes the capture sensor. For these evaluations, the scenario description entails to specify the environment, control mechanisms of this environment, physical layout of the TOE, test equipments and test subjects, their characteristics and their interactions, as well as training and acclimatization methods in addition to the particular software implemented for the aforementioned evaluation. In this case, the software should also measure the environmental conditions and relate them to the test subjects' interactions. Besides, scenario evaluations could require a description of the recruitment of test subjects and how the personal information of these subjects is going to be handled to fulfil privacy requirements.

Evaluators shall examine the test configuration (ATE_FUN.1-3). As biometric performance is very sensitive to environment and human factors, evaluators must check that environmental conditions and test crew composition are being configured and selected as defined in the ST and also that there is consistency between test documentation and guidance documents. Moreover, the biometric system settings (thresholds, timeouts, number of attempts, etc.) shall be set according to the level of effort determined for the evaluation. Evaluators have to consider that in some cases, the configuration could be different in the enrolment and recognition processes, being more restrictive in the first case. Furthermore, if the biometric system is provided with an evaluation operation mode, evaluators must check that the system is configured in this mode for testing.

Evaluators shall examine the test plan to determine if the test execution sequence is proper to obtain the error rates (ATE_FUN.1-4) and consistent with the target application. For technology evaluations, evaluators shall check procedures for processing biometric data of the test corpus. In case of scenario evaluations, they must carefully analyse the number of visits and the activities to be performed in each visit, i.e. training, enrolment, genuine transactions and impostor transactions. In addition, evaluators have to pay attention to aspects such as duration of the visits, composition of the test crew, habituation of test subjects and time between visits and activities, as all of them could affect test subject interactions and modify the quality of the biometric samples.

Evaluators shall examine that test documentation includes the claimed error rates and confidence level to achieve (ATE_FUN.1-5). Also, evaluators have to check that the software implemented for the evaluation saves all the essential information to lately calculate error rates.

Evaluators shall check that the obtained error rates are consistent with the claimed error rates (ATE_FUN.1-6). They have to analyse that all mandatory performance metrics established in ISO/IEC19795 Part 1 and 2 and their corresponding uncertainty have been calculated. Evaluators have to examine that the statistical methods to calculate metrics from the outcome of biometric systems have been applied correctly too.

Evaluators shall report the developer the testing effort by describing, at least, the type of performance evaluation, obtained performance metrics, most relevant details about environment and test subjects, number of interactions (genuine and impostor) and level of effort considered for the evaluation (ATE_FUN.1-7).

7.6.4 ATE_IND: Independent testing

This family specifies how evaluators shall analyse the functional testing performed by developers and how the evaluators themselves shall carry out additional independent tests. Due to the nature of biometric technology and the necessity of using life subjects (scenario evaluations) in many cases, both functional testing and independent testing are two activities that entail too much effort. For these reason, an agreement between developers, evaluators and national schemes should be achieved to deal with biometric performance evaluations as well as functional and independent testing activities in order to reduce cost.

One solution was already proposed in BEM. This document states that biometric performance evaluations must be developed by independent accredited testing facilities for the IT Security Testing Facility (ITSEF). That means that the CC test laboratories use other independent facilities to carry out the biometric performance evaluation. This solution has been adopted for the last certified biometric product [ST'10]. In such case neither specific error rates were claimed in the ST, nor was biometric performance analysed during the evaluation because it has been already assessed by the independent laboratory International Biometric Group (IBG).

Other solution was performed during the certification process of the biometric TOE denominated PalmSecure SDK Version 24 Premium Fujitsu Limited. Its certification report [ST'08] explains that functional and independent test of biometric performance must be carried out together as one unique test because of its large duration. Such test was performed at the developers' facilities in the presence of evaluators. Then, developers provided evaluators with the essential resources and evaluators carried out the same test with independent samples of a small number of test subjects (10% of all samples).

Both solutions are acceptable, but with certain considerations. For the former, accredited laboratories shall have a scope of testing which covers ISO/IEC 19795 Part 1 and Part 2 at least. For the latter, developers and/or evaluators shall reach an agreement about the appropriate test plan. Nevertheless, both shall produce their own test documentation (ATE_IND.1-4 and ATE_IND 2.7). Besides, whoever will carry out tests have sufficient infrastructure and the means to recruit the proper test crew/database. Evaluators always must be able to intervene in order to check that results are not being biased (ATE_IND.2-3 to ATE_IND.2-5).

Moreover, when evaluators want to perform a test subset with a reduced test crew (ATE_IND.1-3 to ATE_IND.1-5 and ATE_IND.2-6 to ATE_IND.2-8) they have to recruit extra people with similar characteristics to the existing test subjects or contact to some of the same test subjects who took part in the functional testing. Developers cannot take part in the

selection of this test population to avoid they choose the proper people to obtain specific results.

Nevertheless, in both cases evaluators shall examine test documentation as it was explained in the Functional test family in addition to check the following during performance testing:

- Correct configuration and installation of the TOE in its operational environment (ATE_IND.1-1, ATE_IND.2-1, ATE_IND.1-2 and ATE_IND.2-2).
- Correct test equipments calibration (ATE_IND.1-1 and ATE_IND.2-2). In case of biometric systems, this requirement includes that test subjects/databases are suitable for the purpose of test. If recruited test subjects are used, procedures for instructing and training them shall be carried out in accordance to test plan specification. If a test corpus is used, procedures for preparing and validating such corpus shall be performed.
- Correct operation of any software implemented to automatically record data (ATE_IND.1-1 and ATE_IND.2-2). In case that test subjects are used, biometric comparison trials shall be executed in accordance to test plan specification (ATE_IND.1-5 and ATE_IND.2-8). In case that a test corpus is used, procedures for processing biometric data shall be carried out in accordance to the test plan specification (ATE_IND.1-5 and ATE_IND.2-8).
- Proper application of the approaches to calculate performance metrics and their uncertainty.
- Analysis of any potential inconsistency between the claimed error rates and the obtained error rates (ATE_IND.1-7 and ATE_IND.2-10).

Furthermore, evaluators have to record and report the documentation describing their effort, all details about the biometric performance testing and the verdict of the activity (ATE_IND.1-6, ATE_IND.1-8, ATE_IND.2-9 and ATE_IND.2-11).

7.7 Considerations for interpreting contours conditions influence on biometric performance in terms of CC

In general, the current version of CC does not cover the evaluation of the operational environment and, as a consequence, the evaluation of contour conditions. Security objectives for the operational environment shall be defined but these are not translated to security functional requirements. It is assumed that the operational environment of the TOE fulfil those security objectives.

However, if the corresponding security objectives for the operational environment are not defined, the influence of environmental conditions and H-B interactions factors are possible vulnerabilities for the security level achieved by biometric system's mechanisms.

This circumstance was already addressed by BEM and the ISO/IEC 19792 standard. While BEM just states that the security evaluation of biometric systems should include an analysis of the dependence of security of environmental factors, the ISO/IEC 19792 standard identifies a hostile environment or the conversion of biometric characteristics as potential vulnerabilities and proposes their analysis as part of the vulnerability assessment. However, neither of them established a specific evaluation methodology to conduct these tasks.

According to CC, the security assurance class that involves vulnerability assessment is AVA. This class is composed by only one family called AVA_VAN. Briefly, the testing activities that entail to satisfy the requirements of this assurance family are the following:

- Identify possible vulnerabilities of the TOE in their operational environment. This process is usually conducted during the application of the rest assurance classes.
- Obtain the attack potential of all the possible vulnerabilities and determine which of them have an attack potential higher that the level established for the EAL selected for the evaluation.
- Conduct penetration tests for checking whether any of potential vulnerabilities are exploitable or not.

Regarding these three testing activities, only two general recommendations are provided:

- The analysis of possible vulnerabilities related to contour conditions influence could be carried out during the application of AGD and ATE classes as follows:
    - Preparative and operational guidelines may be used to identify which environmental conditions and H-B interaction factors may affect biometric system performance in a greater extent.
    - A study of the cases and situations for which biometric performance errors have occurred during the execution of functional tests (ATE_FUN) and/or independent testing (ATE_IND) evaluation activities may help to find potential vulnerabilities as well. It will require the examination of error logs and test operator observations that must reported together with biometric performance results.
- In case of that either environmental conditions or H-B interaction influence shall be determined as potential vulnerabilities, the penetration tests for analyzing them shall be specified based the evaluation methodologies described in Chapter 5 and Chapter 6 respectively. However certain aspects may be considered to reduce the effort that involves the application of such methodologies:
    - The scenario evaluation at the reference evaluation environment/conditions have been already conducted during the application of test procedures for fulfilling ATE_FUN and ATE_IND testing activities.
    - A smaller test crew size may be enough to demonstrate that the particular contour condition affects biometric system performance considerably and

as a consequence, that the potential vulnerability is exploitable. However, it must be justified.

- In case of that other kind of penetration tests easier that the proposed methodologies shall be specified, requirements for generating, controlling, recording and measuring the evaluation conditions could be based on such methodologies.

7.8 Research works developed for defining the proposed guidelines

This section describes different research works that have been conducted for developing and improving the proposed guidelines. This description includes the works which were the starting point and the first studies carried out before the first version of the guidelines. Then, the evolution of these guidelines will be explained including the advances carried out as well as their future.

7.8.1 Preliminary studies and first version of the guidelines

The starting point of the research work performed for this dissertation was a work done by the research group titled "Improvement in Security Evaluation of Biometric Systems" [SAN'06]. This was a first approach to propose an evaluation methodology for improving the application of CC to biometrics. However, this preliminary version was only focused on iris modality and token-based authentication biometric systems. This did not cover all kind of biometric systems. In addition, the evaluation methodology given on it just stated general requirements to analyse certain vulnerabilities. This work did not offer a detailed methodology. In addition, this methodology was not described in terms of CC.

Considering this previous document and its drawbacks, the initial main objective for the research activities on this topic was to develop a rigorous study of CC and biometric systems. It was fundamental to understand biometric systems from a CC point of view. As a result, the work titled "Development of a Protection Profile for Biometric Systems Following ISO/IEC TR 15446" [FER'09] was done. For this work a review of all PPs, STs and documents related to CC and biometric systems was carried out. From those documents and using the guidance provided by the Technical Report ISO/IEC TR 15446 [ISO'09b] it was possible to carefully define the security problem of a general biometric system, identifying which parts of CC need additional guidance in case of testing biometric systems.

Based on these previous works, the first version of the guidelines was developed. This was presented at the International Common Criteria Conference in 2010 with the title "Security Evaluation of Biometric Systems in Common Criteria" [FER'10d]. This version already defined most of the necessary requirements for applying AGD and ATE classes to biometric systems. Nevertheless, those guidelines were not explained in detail, as they had to be related to the corresponding CEM work unit for being more helpful.

7.8.2 Development of the guidelines and its future

Considering the above mentioned first version, a revision of it for improving the aforementioned issues was performed. For this new version the following tasks were done:

- It was fundamental to define the biometric system security functionality and the functions used to implement that functionality according to CC. Most CEM work units have been specified considering the different levels of representation of the TOE: the functional specification and the TOE design. For this reason, a description of both levels for a general biometric system was essential for being more precise when explaining the proposed guidelines.
- Each of the guidelines proposed at the first version was explained in depth. A more comprehensive description providing details and considerations for particular characteristics of certain kinds of biometric systems was given.
- A correspondence between each guideline and its related work unit was carried out. This task was done not only for completing the previous version but also for checking that all work units were addressed.

This version involves the current version of the guidelines which has been presented to both biometric community and CC community. Firstly, it was presented the work titled "Common Criteria and Biometric Performance Testing" [FER'12a] to the biometric community. Then, a revised version titled "Guidelines for Applying AGD and ATE Testing Activities to Biometric Systems" [FER'12b] was presented to the CC community. For this last version, some mistakes were corrected and the most relevant guidelines were illustrated using an example of an automatic handwritten signature verification system.

At last, it is important to mention that the proposed guidelines has been offered to the Spanish national certification body in order to initiate the proper actions to develop a CC formal document for being distributed among developers and evaluators.

7.9 Conclusions

This chapter has presented detailed guidelines for applying CC and CEM to biometric systems. This has been done with the purpose of not only accomplishing the second main objective of this Thesis, i.e. the formalization of the proposed evaluation methodologies in compliance with CC, but also filling the gap that currently exists when interpreting the CEM testing activities in the case of biometric technology.

As a result, specific test actions for conducting biometric performance testing and the analysis of contour conditions influential effects have been defined. Such definition has been based on the already published works in this field, the last version of both, CC documents and the ISO/IEC 19795 multipart standard. Also the evaluation methodologies explained in previous chapters have been used as a basis of this work. Specifically the following aspects have been provided:

- A description of a general biometric system in the context of CC at two levels of abstraction, i.e. the functional specification and the TOE design. This description has been essential to understand this kind of systems in terms of the CC evaluation framework.
- Comprehensive guidelines for applying AGD and ATE classes to biometric systems. For each work unit addressed in CEM regarding both testing activities, specific tasks for carrying out them in case of biometric systems have been described. Also, these guidelines contain particular considerations for certain biometric modalities and kinds of biometric systems.
- An interpretation of the defined environmental and H-B interaction testing methodologies from a CC point of view.

Nevertheless, these guidelines need to be completed in the future, when formal evaluation methodologies that state how to analyse potential vulnerabilities, as well as how to execute penetration tests, will be defined.

The work in this field has provided the following set of publications:

- B. Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno and C. Sanchez-Avila, *Evaluation Methodology for Fake Samples Detection in Biometrics,* 42nd Annual IEEE International Carnahan Conference on Security Technology (ICCST), Prague, 2008 [FER'08b].
- Belen Fernandez-Saavedra, R. Sanchez-Reillo and R. Alonso-Moreno, *Evaluation Methodology Based on CEM for Testing Environmental Influence in Biometric Devices,* International Common Criteria Conference (ICCC)*,* Jeju, 2008 [FER'08a].
- Belen Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno and I. Tomeo-Reyes, *Development of a Protection Profile for Biometric Systems Following ISO/IEC TR 15446,* International Common Criteria Conference (ICCC)*,* Tromso, 2009 [FER'09].
- Belen Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno and I. Tomeo-Reyes, *Security Evaluation of Biometric Systems in Common Criteria,* International Common Criteria Conference (ICCC)*,* Antalya, 2010 [FER'10d].
- Belen Fernandez-Saavedra, R. Sanchez-Reillo, J. Liu-Jimenez and I. Tomeo-Reyes. *Common Criteria and Biometric Performance Testing,* International Biometric Performance Testing Conference (ICBP 2012)*,* Gaithesburg, 2012 [FER'12a].
- Belen Fernandez-Saavedra, R. Sanchez-Reillo, J. Liu-Jimenez and O. Miguel Hurtado, *Guidelines for Applying AGD and ATE Testing Activities to Biometric Systems,* International Common Criteria Conference (ICCC), Paris, 2012 [FER'12b].

Chapter 8

Conclusions and future work lines

This PhD Thesis has proposed new evaluation methodologies for testing biometric systems performance working under specific contour conditions. In particular, two independent evaluation methodologies have been specified for testing the influence of environmental conditions and H-B interaction conditions on biometric system performance respectively. In addition, detailed guidelines have been defined for addressing how to conduct biometric performance and the evaluation of the studied contour conditions in the context of the Common Criteria security evaluations.

This chapter summarizes the main conclusions of the work conducted for the development of these contributions. First, these conclusions will be described separately considering each of the research activities developed. Then, general conclusion about the overall dissertation will be given.

Furthermore, during the development of the research activities carried out for this dissertation, it have been identified certain research lines whose discussion is out of the scope of it. Nevertheless, this chapter also describes these open research lines that should be addressed by future works.

8.1 Conclusions

After the whole description of the work conducted in this PhD Thesis, this section describes the main conclusions obtained.

8.1.1 Contour conditions evaluation methodologies

Regarding the first objective of this Thesis, two evaluation methodologies have been defined for analysing two of the major factors that have been traditionally claimed as influential factors, i.e. ambient conditions and human-biometric system interaction conditions respectively.

Firstly, a complete methodology for evaluating the influence of ambient conditions on the performance of biometric systems has been specified. This methodology presents the following characteristics:

- It has been defined following the principles and requirements of ISO/IEC 19795 multipart standard.
- The defined methodology is general for covering different recognition mechanisms (e.g. verification, open-set and closed-set identification) and biometric modality independent. It is also independent on the specific technology used for acquiring the biometric samples.
- Defines the way that different ambient conditions parameters shall be selected in order to carry on the evaluation.
- Defines equipment and tools needed to generate, control, measure and record the particular evaluation conditions.
- Defines test procedures for planning, conducting and reporting environmental testing of biometric systems as part of a biometric performance scenario evaluation.
- Defines the procedures to compare the performance obtained at specific ambient conditions, with a baseline performance used as a reference.
- Defines the measures needed for quantifying the level of influence of each set of ambient conditions.

This methodology has been offered and accepted into ISO/IEC JTC1/SC37, in order to become an International Standard, which will be published in the near future under the number ISO/IEC 29197.

Likewise, regarding the evaluation of the influence of human factors on biometric system performance, the methodology developed has accomplished the following targets:

- It has been defined following the principles and requirements of ISO/IEC 19795 multipart standard.
- It is also based on previous works such as the HBSI evaluation framework and usability studies conducted by NIST.

- The defined methodology is general for covering different recognition mechanisms (e.g. verification, open-set and closed-set identification) and biometric modality independent. It is also independent on the specific technology used for acquiring the biometric samples.
- It has updated HBSI evaluation framework proposing the fundamental parameters needed for conducting a comprehensive evaluation of the influence of human factors.
- Defines the way the different human-related factor parameters shall be selected in order to carry on the evaluation.
- Defines test procedures for planning, conducting and reporting H-B interaction testing of biometric systems as part of a biometric performance scenario evaluation.
- Defines the measurements needed for quantifying the level of influence of the target conditions.
- Defines the procedures to compare the performance obtained in target conditions, with a baseline performance used as a reference.

The methodology has been written in a way to allow a future offer to the standardization bodies (either in ISO/IEC JTC1/SC37 or in CEN TC224 WG18), so as to be adopted internationally.

8.1.2 Guidelines for Common Criteria evaluation of biometric systems

Regarding the second objective and based on previous works, such as BTSE, BEM and ISO/IEC 19792, specific guidelines needed to address the evaluation of biometric systems in the context of Common Criteria have been defined. In order to reach this goal, the following objectives have been achieved:

- The methodology has been updated to the last published versions of Common Criteria documents and the ISO/IEC 19795 multipart standard.
- Guidelines have been created which are general for all kinds of biometric systems (i.e. verification and identification systems) and modality-independent.
- Biometric systems and their modules have been defined in terms of Common Criteria.
- It has been defined how to interpret CEM testing activities when being applied to the analysis of the performance of biometric systems.
- Specific evaluation tasks have been defined for AGD and ATE assurance classes for conducting biometric performance evaluation.
- It has been described how to measure the impact of ambient conditions and human factors in the scope of Common Criteria.

The proposed guidelines have been offered to the Spanish National Certification Body in order to initiate the proper actions to develop a formal CC supporting document.

8.1.3 General conclusions

As an overall conclusion of the whole work done in this Thesis, it can be highlighted that biometric technologies and products are not currently evaluated in a comprehensive way, due to the lack of the existence of evaluation methodologies, and the cost (both in time and in money) for carrying all the needed testing processes to achieve a real generalized assessment of the biometric system. With the methodologies developed in this Thesis, it is expected to minimize such inconvenience.

Regarding security evaluation, the use of Common Criteria may not be the most suitable approach, due to the initial design of Common Criteria, which do not easily consider some changing contour conditions, such as ambient conditions and human factors. Therefore a specific evaluation framework should be established for the assessment of the security of biometric systems. In such a new framework, not only the above mentioned parameters shall be included, but also topics which are significant for biometrics such as liveness detection, spoofing resistance, etc.

8.2 Future works

As it has been explained in the previous chapters, there are certain research works that have been impossible to be covered as part of this PhD Thesis. In addition, from all the activities carried out, several open research lines are available for the interest of the scientific and industrial community. Some of these research works and lines are the following:

- To conduct more evaluations on biometric products of different modalities and which use different kinds of biometric capture devices based on the methodologies developed in this Thesis. Furthermore, test their level of interoperability and repeatability when different evaluation laboratories are considered.
- To complete the evaluation methodology for environmental testing including further environmental parameters in addition to those in the scope of this work, such as vibration or atmospheric pressure.
- To complete the evaluation methodology for H-B interaction testing to all parameters that have been defined in section 6.2.2.
- As it has been shown, the evaluation methodologies for the contour conditions of the biometric systems have many points in common. Therefore it is considered as feasible to develop a general evaluation methodology based on the two evaluation methodologies that have been defined in this PhD Thesis, providing it to standardization committees.
- To fulfil Common Criteria guidelines for the analysis of all kinds of vulnerabilities in biometric systems. As some vulnerabilities are heavily dependent on the modality chosen, different guidelines will have to be defined for each of the vulnerabilities present in each biometric modality.

- To evolve the work here performed with the potential future evolutions of Common Criteria, such as the ones known as Common Criteria Light, or as Collaborative Protection Profiles.

References

[ANSI'11] ANSI, *ANSI/INCITS 473 - 2011, Information Technology -- Conformance Testing Methodology Standard for Patron Formats Conforming to INCITS 398-2008, Information Technology -- Common Biometric Exchange Formats Framework (CBEFF)*, ANSI/INCITS 473-2011, 2011.

[ASD'99] Analytical Spectral Devices Inc., *ASD Technical Guide 3rd Ed.*, 1999.

[ATHOS'05] Athos Origin, *UK Passport Service: Biometrics Enrolment Trial*, 2005.

[AUT'07] Authenti-Corp, *Iris Recognition Study 2006 (IRIS06), Standards-Based Performance and User Cooperation Studies of Commercial Iris Recognition Products*, Authenti-Corp, 2007.

[BEM'02] Common criteria Biometric Evaluation Methodology Working Group, *Biometric Evaluation Methodology Supplement*, 2002.

[BEV'10] J. R. Beveridge, D. S. Bolme, B. A. Draper, et al., *Quantifying how lighting and focus affect face recognition performance,* IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2010.

[BTSE'01] Electronic Warfare Associates-Canada Ltd., *Biometric Technology Security Evaluation under the Common Criteria*, 2001.

[BWG'00] Biometrics Working Group, *Best Practices in Testing and Reporting Performance of Biometric Devices Version 1.0*, 2000.

[CC] Common Criteria (CC), *The Common Criteria Portal* http://www.commoncriteriaportal.org/.

[CC-1'96] Common Criteria, *Common Criteria for information Technology Security, Version 1.0*, 1996.

[CC-1'99] Common Criteria, *Common Criteria for information Technology Security Evaluation, Version 2.1*, 1999.

[CC-1'12] Common Criteria, *Common Criteria for information Technology Security Evaluation, Version 3.1, Revision 4*, 2012.

[CC-2'12] Common Criteria, *Common Criteria for information Technology Security Evaluation, Version 3.1, Revision 4*, 2012.

[CC-3'12] Common Criteria, *Common Criteria for information Technology Security Evaluation, Version 3.1, Revision 4*, 2012.

[CCN'08] Centro Criptológico Nacional (CCN), *Characterizing Attacks to Fingerprint Verification Mechanisms, Version 1.0*, 2008.

[CEM'12] Common Criteria, *Common Methodology for Information Technology Security Evaluation, Version 3.1, Revision 4*, 2012.

[CEN'11] European Committee for Standardization, *CEN EN 12464-1: Ligh and lighting - Lighting of work places - Part 1 : Indoor work places*, CEN EN 12464-1, 2011.

[CESG3'89] Communications-Electronics Security Group (CSEG), *UK Systems Security Confidence Levels, CESG Computer Security Memorandum No. 3*, 1989.

[COV'03] Lynne Coventry, Antonella De Angeli, and Graham Johnson, *Biometric Verification at a Self Service Interface,* British Ergonomic Society Conference, 2003.

[CTCPEC'93] Government of Canada, *The Canadian Trusted Computer Product Evaluation Criteria (CTCPEC)*, 1993.

[DEF'06] Ministry of Defence; Defence Standard 00-35 Issue 4, *Environmental Handbook for Defence Material - Part 3: Environmental Test Methods* DEF STAN 00-35 Part 3 Issue 4, 2006.

[DIR'03] The European Parliament and the Council of the European Union, *Directive 2003/10/EC of the European Parliament and of the Council of 6 February 2003 on the minimum health and safety requirements regarding the exposure of*

workers to the risks arising from physical agents (noise), Official Journal of the European Union, 2003.

[DoD'73] Department of Defense, *DoD 5200.28-M - ADP Security Manual, Techniques and Procedures for Implementing, Deactivating, Testing, and Evaluating -Secure Resource-Sharing ADP System*, 1973.

[DOD'98] G. Doddington, W. Liggett, A. Martin, et al., *Sheep, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation,* Int'l Conf. Spoken Language Processing (ICSLP)*,* Sydney, 1998.

[DOD'00] George R. Doddington, Mark A. Przybocki, Alvin F. Martin, et al., *The NIST speaker recognition evaluation – Overview, methodology, systems, results, perspective,* Speech Communication, 31(2–3), p. 225-254, 2000.

[DTIEC'89] Department of Trade and Industry, *DTI Commercial Computer Security Centre Evaluation Levels Manual, V22*, 1989.

[DUA'01] Duane M. Blackburn, Mike Bone, and P. Jonathon Phillips, *Facial Recognition Vendor Test 2000: Evaluation Report*, 2001.

[DUN'09] Ted Dunstone and Yager Neil, *Biometric Systems and Data Analysis. Design, Evaluation, and Data Mining*, Springer, 2009.

[ELL'10] S. Elliott and E. Kukula, *A Definitional Framework for the Human-Biometric Sensor Interaction Model*, Proceedings of SPIE - The International Society for Optical Engineering ed, Society of Photo-Optical Instrumentation Engineers, 7667, 2010.

[ELL'12] Stephen Elliott, *Evolution of the Human Biometric Sensor Interaction,* International Biometric Performance Testing Conference (ICBP 2012), 2012.

[ERB'12] M. Erbilek and M. Fairhurst, *Framework for managing ageing effects in signature biometrics,* Biometrics, IET, 1(2), p. 136-147, 2012.

[FAI'05] M. C. Fairhurst and C. McIntosh, *Assessing image characteristics for user feedback in biometric fingerprint identity verification tasks,* IEEE International Conference on Visual Information Engineering (VIE 2005), 2005.

[FAI'11] M. Fairhurst and M. Erbilek, *Analysis of physical ageing effects in iris biometrics,* Computer Vision, IET, 5(6), p. 358-366, 2011.

[FCITS'92a] National Institute of Standards and Technology and National Security Agency, *Federal Criteria for Information Technology Security, Volume II: Registry of Protection Profile Version 1.0*, 1992a.

[FCITS'92b] National Institute of Standards and Technology and National Security Agency, *Federal Criteria for Information Technology Security, Volume I: Protection Profile Development Version 1.0*, 1992b.

[FER'08a] Belen Fernandez-Saavedra, Raul Sanchez-Reillo, and Raul Alonso-Moreno, *Evaluation Methodology Based on CEM for Testing Environmental Influence in Biometric Devices,* International Common Criteria Conference (ICCC), 2008a.

[FER'08b] B. Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno, et al., *Evaluation methodology for fake samples detection in biometrics,* 42nd Annual IEEE International Carnahan Conference on Security Technology (ICCST), 2008b.

[FER'08c] B. Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno, et al., *Evaluation methodology for analyzing environment influence in biometrics,* 10th International Conference on Control, Automation, Robotics and Vision (ICARCV) 2008c.

[FER'09] Belen Fernandez-Saavedra, Raul Sanchez-Reillo, Raul Alonso-Moreno, et al., *Development of a Protection Profile for Biometric Systems Following ISO/IEC TR 15446,* International Common Criteria Conference (ICCC), 2009.

[FER'10a] B. Fernandez-Saavedra, R. Alonso-Moreno, A. Mendaza-Ormaza, et al., *Usability Evaluation of Fingerprint Based Access Control Systems,* 2010 Sixth International

Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010a.

[FER'10b] B. Fernandez-Saavedra, R. Alonso-Moreno, J. Uriarte-Antonio, et al., *Evaluation methodology for analyzing usability factors in biometrics,* IEEE Transactions on Aerospace and Electronic Systems Magazine, 25(8), p. 20-31, 2010b.

[FER'10c] B. Fernandez-Saavedra, F. J. Diez-Jimeno, R. Sanchez-Reillo, et al., *Establishment of baseline performance for "end to end" biometric system evaluations,* IEEE International Carnahan Conference on Security Technology (ICCST), 2010c.

[FER'10d] Belen Fernandez-Saavedra, Raul Sanchez-Reillo, Raul Alonso-Moreno, et al., *Security Evaluation of Biometric Systems in Common Criteria,* International Common Criteria Conference (ICCC), 2010d.

[FER'10e] Belen Fernandez-Saavedra, Raul Sanchez-Reillo, Raul Alonso-Moreno, et al., *Environmental Testing Methodology in Biometrics,* International Biometric Performance Testing Conference (ICBP 2010), 2010e.

[FER'11] B. Fernandez-Saavedra, I. Tomeo-Reyes, Francisco J. Diez-Jimeno., et al., *Operational and Security Evaluation of Authentication Systems in Critical Infrastructures,* 4th International Conference on Experiments/Process/System Modeling/Simulation/Optimization, 2011.

[FER'12a] Belen Fernandez-Saavedra, Raul Sanchez-Reillo, Judith Liu-Jimenez, et al., *Common Criteria and Biometric Performance Testing,* International Biometric Performance Testing Conference (ICBP 2012), 2012a.

[FER'12b] Belen Fernandez-Saavedra, Raul Sanchez-Reillo, Judith Liu-Jimenez, et al., *Guidelines for Applying AGD and ATE Testing Activities to Biometric Systems,* International Common Criteria Conference (ICCC), 2012b.

[FERET'11] National Institute of Standards and Technology (NIST), *Face Recognition Technology (FERET)*, 2011.

[GUEST'06] Richard Guest, *Age dependency in handwritten dynamic signature verification systems,* Pattern Recognition Letters, 27(10), p. 1098-1104, 2006.

[HAB'76] Wolf Haberman and Adolph Fejfar, *Automatic Identification of Personnel through Speaker and Signature Verification - System Description and Testing,* Carnahan Conference on Crime Countermeasures, 1976.

[HAZ'06] Hazel Lacohée, Stephen Crane, and Andy Phippen, *Trustguide: Final Report*, 2006.

[HEN'10] Olaf Henniger, Dirk Scheuermann, and Thomas Kniess, *On security evaluation of fingerprint recognition systems,* 1st International Biometric Performance Conference, Gaithersburg, 2010.

[IBG'02] International Biometric Group (IBG), *Comparative Biometric Testing - Test Plan 2.1*, 2002.

[IBG'03] International Biometric Group (IBG), *Lessons Learned from Comparative Biometric Testing*, 2003.

[IBG'06] International Biometric Group (IBG), *Comparative Biometric Testing - Round 6 Public Report*, 2006.

[IBG'09] International Biometric Group (IBG), *Comparative Biometric Testing - Round 7 Public Report*, 2009.

[IBG'12] International Biometric Group (IBG), *Comparative Biometric Testing* 2012. http://www.ibgweb.com/about/case-studies/comparative-biometric-testing.

[IEA'00] International Ergonomics Association (IEA), *IEA Ergonomics human centered design*, 2000. http://www.iea.cc/01_what/What%20is%20Ergonomics.html.

[IEC'12] IEC, *International Electrotechnical Commission*, 2012. http://www.iec.ch/.

[IEEE'09a] IEEE Certified Biometrics Professional (CBP), *Module 2 Biometric Modalities*, 2009a.

[IEEE'09b] IEEE Certified Biometrics Professional (CBP), *Module 4 Biometrics Standards*, 2009b.

[IEEE'09c] IEEE Certified Biometrics Professional (CBP), *Module 1 Biometrics Fundamentals*, 2009c.

[IEEE'09d] IEEE Certified Biometrics Professional (CBP), *Module 3 Biometric System Design and Evaluation*, 2009d.

[ISO'98] International Organization for Standardization, *ISO 9241: Ergonomic requirements of human system interaction for office work with visual display terminals (VDTs) – Part 11: Guidance on usability*, ISO 9241-11:1998, 1998.

[ISO'06a] International Organization for Standardization, *ISO/IEC 19784-1, Information technology -- Biometric application programming interface -- Part 1: BioAPI specification*, ISO/IEC 19784-1:2006, 2006a.

[ISO'06b] International Organization for Standardization, *ISO/IEC 19795-1, Information technology -- Biometric performance testing and reporting -- Part 1: Principles and framework*, ISO/IEC 19795-1:2006, 2006b.

[ISO'06c] International Organization for Standardization, *ISO/IEC 19785-1, Information technology -- Common Biometric Exchange Formats Framework -- Part 1: Data element specification*, ISO/IEC 19785-1:2006, 2006c.

[ISO'07a] International Organization for Standardization, *ISO/IEC 19795-2, Information technology -- Biometric performance testing and reporting -- Part 2: Testing methodologies for technology and scenario evaluation*, ISO/IEC 19795-2:2007, 2007a.

[ISO'07b] International Organization for Standardization, *ISO/IEC TR 24741, Information technology -- Biometrics tutorial*, ISO/IEC TR 24741:2007, 2007b.

[ISO'07c] International Organization for Standardization, *ISO/IEC TR 19795-3, Information technology -- Biometric performance testing and reporting -- Part 3: Modality-specific testing*, ISO/IEC TR 19795-3:2007, 2007c.

[ISO'07d] International Organization for Standardization, *ISO/IEC 24709-1, Information technology -- Conformance testing for the biometric application programming interface (BioAPI) -- Part 1: Methods and procedures*, ISO/IEC 24709-1:2007, 2007d.

[ISO'08] International Organization for Standardization, *ISO/IEC 19795-4, Information technology -- Biometric performance testing and reporting -- Part 4: Interoperability performance testing*, ISO/IEC 19795-4:2008, 2008.

[ISO'09a] International Organization for Standardization, *ISO/IEC 19792, Information technology -- Security techniques -- Security evaluation of biometrics*, ISO/IEC 19792:2009, 2009a.

[ISO'09b] International Organization for Standardization, *ISO/IEC TR 15446, Information technology -- Security techniques -- Guide for the production of Protection Profiles and Security Targets*, ISO/IEC TR 15446:2009, 2009b.

[ISO'09c] International Organization for Standardization, *ISO/IEC 29109-1, Information technology -- Conformance testing methodology for biometric data interchange formats defined in ISO/IEC 19794 -- Part 1: Generalized conformance testing methodology*, ISO/IEC 29109-1:2009, 2009c.

[ISO'09d] International Organization for Standardization, *ISO/IEC 29794-1, Information technology -- Biometric sample quality -- Part 1: Framework*, ISO/IEC 29794-1:2009, 2009d.

[ISO'10a] International Organization for Standardization, *ISO/IEC Standing Document 11 (SD11) -- Part 1: Harmonization document*, ISO/IEC SD11 2010a.

[ISO'10b] International Organization for Standardization, *ISO/IEC 29794-5, Information technology -- Biometric sample quality -- Part 4: Face image data*, 2010b.

[ISO'10c] International Organization for Standardization, *ISO/IEC 29794-4, Information technology -- Biometric sample quality -- Part 4: Finger image data*, 2010c.

[ISO'11a] International Organization for Standardization, *ISO/IEC 19794-1, Information technology -- Biometric data interchange formats -- Part 1: Framework*, 2011a.

[ISO'11b] International Organization for Standardization, *ISO/IEC 19795-7, Information technology -- Biometric performance testing and reporting -- Part 7: Testing of on-card biometric comparison algorithms*, ISO/IEC 19795-7:2011, 2011b.

[ISO'11c] International Organization for Standardization, *ISO/IEC 19795-5, Information technology -- Biometric performance testing and reporting -- Part 5: Access control scenario and grading scheme*, ISO/IEC 19795-5:2011, 2011c.

[ISO'12a] ISO, *International Organization for Standardization*, 2012a. http://www.iso.org/iso/home.html.

[ISO'12b] International Organization for Standardization, *ISO/IEC CD 29794-6, Information technology -- Biometric sample quality -- Part 4: Iris image data*, 2012b.

[ISO'12c] International Organization for Standardization, *ISO/IEC WD 30107, Information technology -- Antispoofing and liveness detection techniques*, ISO/IEC WD 30107, 2012c.

[ISO'12d] International Organization for Standardization, *ISO/IEC 2382-37, Information technology -- Vocabulary -- Part 37: Biometrics*, 2012d.

[ISO'12e] International Organization for Standardization, *ISO/IEC 19795-6, Information technology -- Biometric performance testing and reporting -- Part 6: Testing methodologies for operational evaluation*, ISO/IEC 19795-6, 2012e.

[ISO'12f] International Organization for Standardization, *ISO/IEC CD 29197, Information technology -- Evaluation methodology for environmental influence in biometric system performance*, ISO/IEC CD 29197, 2012f.

[ITSEC'91] Commission of the European Communities, *Information Technology Security Evaluation Criteria (ITSEC)*, 1991.

[JAI'07] Anil K. Jain, Patrick Flynn, and Arun A. Ross, *Handbook of Biometrics*, Springer-Verlag New York Inc., 2007.

[JAIN'98] Anil K. Jain, Ruud Bolle, and Sharath Pankanti, *Biometrics, Personal Identification in Networked Society*, Kluwer Academic Publishers, 1998.

[JON'96] P. Jonathon Phillips, Patrick J. Rauss, and Sandor Z. Der, *FERET (Face Recognition Technology) Recognition Algorithm Development and Test Results*, Army Research Laboratory, 1996.

[JON'00] P. Jonathon Phillips, Alvin Martin, C.L. Wilson, et al., *An Introduction to Evaluating Biometric Systems,* National Institute of Standards and Technology (NIST), 2000.

[KANG'03] Hyosup Kang, Bongku Lee, Hakil Kim, et al., *A Study on Performance Evaluation of Fingerprint Sensors*, Springer Berlin Heidelberg, 2003.

[KIM'03] Hale Kim, *Evaluation of Fingerprint Readers: Environmental factors, Human Factors, & Liveness Detecting Capability*, 2003, http://www.biometrics.org/bc2004/CD/PDF_PROCEEDINGS/Microsoft%20PowerPoint%20-%20Presentation%20of%20HaleKim%20-%20v2.1.ppt%20%5B.pdf.

[KUK'04] E. P. Kukula, S. J. Elliott, R. Waupotitsch, et al., *Effects of illumination changes on the performance of Geometrix FaceVision 3D FRS,* 38th Annual 2004 International Carnahan Conference on Security Technology, 2004.

[KUK'06] E. P. Kukula and S. J. Elliott, *Implementing Ergonomic Principles in a Biometric System: A Look at the Human Biometric Sensor Interaction (HBSI),* 40th Annual IEEE International Carnahan Conferences Security Technology, 2006.

[KUK'07] E. P. Kukula, S. J. Elliott, B. P. Gresock, et al., *Defining Habituation using Hand Geometry,* IEEE Workshop on Automatic Identification Advanced Technologies, 2007.

[KUK'08] E. P. Kukula, *Design and Evaluation of the Human-Biometric Sensor Interaction Method*, Center for Education and Research in Information Assurance and Security, Purdue University, West Lafayette, 2008.

[KUK'10] E. P. Kukula, M. J. Sutton, and S. J. Elliott, *The Human Biometric-Sensor Interaction Evaluation Method: Biometric Performance and Usability Measurements,* IEEE Transactions on Instrumentation and Measurement, 59(4), p. 784-791, 2010.

[LI'09] Stan Z. Li and Anil K. Jain, *Encyclopedia of Biometrics*, Springer Publishing Company Inc. , 2009.

[MAG'11] S.A. Magnet, *When Biometrics Fail: Gender, Race, and the Technology of Identity*, Duke University Press, 2011.

[MAI'02] D. Maio, D. Maltoni, R. Cappelli, et al., *FVC2000: fingerprint verification competition,* IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(3), p. 402-412, 2002.

[MAL'09] Davide Maltoni, Dario Maio, Anil K. Jain, et al., *Handbook of Fingerprint Recognition*, Springer Publishing Company Inc., 2009.

[MAN'01] Tony Mansfield, Gavin Kelly, David Chandler, et al., *Biometric Product Testing Final Report Issue 1.0*, National Physical Laboratory (NPL), 2001.

[MAN'02] A. J. Mansfield and J.L. Wayman, *Best Practices in Testing and Reporting Performance of Biometric Devices Version 2.01*, National Physical Laboratory, 2002.

[McG'11] McGraw-Hill, *The McGraw-Hill Dictionary of Scientific and Technical Terms*, McGraw Hill, 2011.

[MER'12] Meryem Erbilek and Michael Fairhurst, *A methodological framework for investigating age factors on the performance of biometric systems,* Proceedings of the on Multimedia and security, ACM, 2012.

[MIC'08] Michael D. Frick, Shimon k. Modi, Stephen Elliot, et al., *Impact of Gender on Fingerprint Recognition Systems,* 5th International Conference on Information Technology and Applications, 2008.

[MIL'08] United States Military Standards, *Department of Defense: Test Method Standard for Environmental Engineering Considerations and Laboratory Test*, MIL-STD-810 G, 2008.

[MOD'06] Shimon K. Modi and Stephen J. Elliott, *Impact of Image Quality on Performance: Comparison of Young and Elderly Fingerprints,* Springer Verlag, 2006.

[MOR'10] Aythamin Morales, Miguel A. Ferrer, Carlos M. Travieso, et al., *About user acceptance in hand, face and signature biometric systems,* Jornadas de Reconocimiento Biometrico de Personas, 2010.

[MUN'12] Axel Munde, *Biometrics and IT Security,* 2th International Biometric Performance Conference, Gaithersburg, 2012.

[NATO'94] North Atlantic Treaty Organization, *NATO STANAG 4370, AECTP 200, Category 230*, 1994.

[NIST'06a] M. Theofanos, R. Micheals, J. Scholtz, et al., *Does Habituation Affect Fingerprint Quality?*, National Institute of Standards and Technology (NIST), 2006a.

[NIST'06b] M. Theofanos, S. Orandi, R. Micheals, et al., *NISTIR 7382; Effects of Scanner Height on Fingerprint Capture*, National Institute of Standards and Technology (NIST), 2006b.

[NIST'06c] M. Theofanos, B. Stanton, S. Orandi, et al., *NISTIR 7403; Usability Testing of Ten-Print fingerprint Capture*, National Institute of Standards and Technology (NIST), 2006c.

[NIST'06d] National Institute of Standards and Technology (NIST), *Biometrics and Usability; Efficiency, Effectiveness and User Satisfaction*, National Institute of Standards and Technology (NIST), 2006d.

[NIST'08] National Institute of Standards and Technology (NIST), *Usability & Biometrics; Ensuring Successful Biometric Systems*, NIST, 2008.

[NSTC'06a] National Science and Technology Council (NSTC), *Privacy & Biometrics. Building a Conceptual Foundation*, 2006a.

[NSTC'06b] National Science and Technology Council (NSTC), *Biometrics Glossary*, 2006b.

[NSTC'06c] National Science and Technology Council (NSTC), *Biometrics History*, 2006c.

[ORC'02] ORC International, *Public Attitudes Toward the Uses of Biometric Identification Technologies by Government and the Private Sector*, 2002.

[OXF'10] Oxford University Press, *Oxford Dictionary of English*, Oxford University Press, 2010.

[PP'05] M. Krechel and N. Tekampe, *Protection Profile for Biometric Verification Mechanisms*, Bundesamt für sicherheit in der Informationstechnik (BSI), 2005.

[PP'08] N. Tekampe and B. Leidner, *Protection Profile for Biometric Verification Mechanisms*, Bundesamt für sicherheit in der Informationstechnik (BSI), 2008.

[PRO'11] H. Proenca, *Quality Assessment of Degraded Iris Images Acquired in the Visible Wavelength,* IEEE Transactions on Information Forensics and Security, 6(1), p. 82-95, 2011.

[RAND'80] R. Stockton Gaines, William Lisowski, James Press, et al., *Authentication by Keystroke Timing: Some Preliminary Results*, The Rand Corporation, 1980.

[RON'03] Ron Sutton, *Status of US and International Biometric Testing and Reporting Standardization Efforts*, Biometric Consortium Conference (BCC), 2003.

[SAN'06] Sanchez-Reillo Raul, Liu-Jimenez Judith, G. Lorenz Michael, et al., *Improvement in Security Evaluation of Biometric Systems,* 40th Annual IEEE International Carnahan Conferences Security Technology (ICCST), 2006.

[SAN'09] R. Sanchez-Reillo, B. Fernandez-Saavedra, J. Liu-Jimenez, et al., *Changes to vascular biometric system security & performance,* IEEE Transaction on Aerospace and Electronic Systems Magazine, 24(6), p. 4-14, 2009.

[SCSSI'89] Service Central de la Sécurité des Systèmes d'Information, *Catalogue de Critères Destinés à évaluer le Degré de Confiance des Systèmes d'Information*, 1989.

[SIC'05] N. C. Sickler and S. J. Elliott, *An evaluation of fingerprint image quality across an elderly population vis-a-vis an 18-25 year old population,* 39th Annual 2005 International Carnahan Conference on Security Technology (ICCST), 2005.

[SRE'12] National Institute of Standards and Technology (NIST), *Speaker Recognition Evaluation*, 2012. http://www.nist.gov/itl/iad/mig/sre.cfm.

[ST'08] Fujitsu, *Security Target for Palm Secure*, Bundesamt für sicherheit in der Informationstechnik (BSI), 2008.

[ST'10] Authenware, *Authentest Server v1.2.6, Declaración de seguridad*, Ministerios de Defensa, Centro Nacional de Inteligencia, Centro Criptológico Nacional, 2010.

[TCSEC'83] US Government, *CSC-STD-001-83 - Department of Defense Trusted Computer System Evaluation Criteria (TCSEC)*, 1983.

[TCSEC'85] Department of Defense, *DoD 5200.28-STD - Department of Defense Trusted Computer System Evaluation Criteria (TCSEC)*, 1985.

[UKBWG'12] UK Biometric Working Group (CESG), *MS07 Health & Safety*, 2012.

[WAY'97] James Wayman, *Biometric Identification Standards Research, Final Report*, Collage of Enfineering, San Jose State University, 1997.

[WAY'00] James L. Wayman, *Technical Testing and Evaluation of Biometric Identification*, National Biometric Test Center Collected Works 1997-2000, 2000.

[WAY'04] J. L. Wayman, A. K. Jain, D. Maltoni, et al., *Biometric Systems: Technology, Design and Performance Evaluation*, Springer-Verlag New York Inc., 2004.

[ZSIEC'89] Bundesamt für Sicherheit in der Informationstechnik (BSI), *Criteria for the Evaluation of Trustworthiness of Information Technology (IT) Systems*, 1989.

[ZWI'00] A. Zwiesele, A. Munde, C. Busch, et al., *BioIS study. Comparative study of biometric identification systems,* IEEE 34th Annual 2000 International Carnahan Conference on Security Technology, 2000.