TECHNICAL REPORT

ISO/IEC TR 19795-3

First edition
2007-12-15

Information technology — Biometric performance testing and reporting —

Part 3:
Modality-specific testing

Technologies de l'information — Essais et rapports de performances biométriques —

Partie 3: Essais de modalité spécifique

Reference number
ISO/IEC TR 19795-3:2007(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword ... iv
Introduction ... v
1 Scope ... 1
2 Normative references ... 1
3 Terms and definitions ... 1
4 Symbols and abbreviated terms ... 2
5 Modality-specific test design ... 2
5.1 Development flow ... 2
5.2 Modality-specific factors that may impact performance (Step 1) ... 3
5.3 Modality-specific policies pertaining to Test Subjects (Step 2) ... 3
5.4 Modality-specific policies pertaining to data collection (Step 3) ... 5
5.5 Modality-specific policies in impostor transactions (Step 4) ... 7
5.6 Modality-specific reporting policies (Step 5) ... 9
6 Evaluation of modality-specific influential factors ... 10
6.1 Robustness tests ... 10
6.2 Base example: face ... 11
6.3 Other examples: fingerprint, iris, vein, and voice ... 13
7 Test design principles on “new” modalities ... 18
Bibliography ... 19

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, the joint technical committee may propose the publication of a Technical Report of one of the following types:

— type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;

— type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;

— type 3, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example).

Technical Reports of types 1 and 2 are subject to review within three years of publication, to decide whether they can be transformed into International Standards. Technical Reports of type 3 do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 19795-3, which is a Technical Report of type [1/2/3], was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

ISO/IEC TR 19795 consists of the following parts, under the general title *Information technology — Biometric performance testing and reporting*:

— *Part 1: Principles and framework*

— *Part 2: Testing methodologies for technology and scenario evaluation*

— *Part 3: Modality-specific testing* [Technical Report]

Performance and interoperability testing of data interchange formats will form the subject of a future Part 4.

Introduction

In biometric performance testing and reporting, careful consideration should be given to the characteristic differences of each modality (fingerprint, face, iris, etc.). These differences naturally require variations within the general methodology defined in ISO/IEC 19795-1.

Since there are many modality-dependent variations, it is desirable that each modality be specified with specific testing and reporting methods. This ensures that not only specialists of biometrics but also non-specialists can carry out reasonably accurate testing.

The characteristics that affect evaluation methodology can at least be discussed from the following viewpoints:

— characteristics of users, including the definition of impostors;

— restrictions that come from practical situations in which that biometric modality is used in applications.

Characteristics of impostors:

There are two factors to consider regarding the definition of impostors: (1) multiple biometric data from one person, and (2) impostor attempts for behaviour-based modalities, such as voice or signature.

For modalities in which multiple biometric data can be collected from one person, e.g. finger (ten fingerprints from one person) and iris (two iris-images from one person), a rule for permitting or prohibiting use of these data as impostor attempts needs to be clearly defined.

In the case of behaviour-based modalities, testing results regarding impostor attempts [false match rate (FMR) or false acceptance rate (FAR)] can be influenced depending on whether (or how much) an impostor tries to imitate an authorized user's behaviour or not. For instance, the case in which an impostor physically traces an authorized user's signature that the impostor obtained differs significantly in FMR or FAR from the case where the impostor only looks at the signature and imitates it. For these modalities, a criterion regarding impostor attempts needs to be defined.

Characteristics of modality specific to applications

In general, almost all modalities of biometrics are used for user authentication. However, some modalities are expected to be used in different classes of applications; for example, face-based identification is widely used in surveillance applications. While we can expect a user's cooperation in the former, we cannot expect it in the latter case. Thus variation of testing methodologies needs to be considered depending on the way the modality is used in real applications.

These restrictions can be divided into two classifications:

— factors relating to users, such as facial expressions that affect the countenance of the face, wearing eye-glasses or contact lenses for the iris;

— factors relating to external environments that are uncontrollable by the algorithm or system, such as illumination change for face or background noise for voice.

These factors naturally affect the performance, and the types and number of factors are different in each modality. These modality-dependent variations need to be considered in testing and reporting. In addition, a concept of robustness testing needs to be introduced to evaluate the sensitivity or robustness of the technology toward environmental factors, in case the variation of the factors strongly influences the observed performance.

This part of ISO/IEC 19795 is intended to describe the variations in methodologies relating to these modality-dependent characteristics. It presents and defines methods for determining, given a specific biometric modality, how to develop a technical performance test.

This part of ISO/IEC 19795 has been developed based on six technical reports [1-6], which are the outputs of standardization activities in Japan for biometric testing and reporting, and have been published by the Japanese Standards Association as JIS-TRs (Japanese Industrial Standards Technical Reports). They have been prepared through discussions by experts of respective modalities, and have extensive considerations particular to different procedures specific to each modality. They are intended to define detailed procedures for testing, including specifications of a test database and how to collect data. These documents can be used as reference when designing specific evaluation procedures.

Information technology — Biometric performance testing and reporting —

Part 3: Modality-specific testing

1 Scope

In biometric performance testing and reporting, careful consideration needs to be given to the characteristic differences of each modality (fingerprint, face, iris, etc.). These differences naturally require variations within the general methodology defined in ISO/IEC 19795-1.

This part of ISO/IEC 19795 describes the methodologies relating to these modality-dependent variations. It presents and defines methods for determining, given a specific biometric modality, how to develop a technical performance test.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 19795-1:2006, *Information technology — Biometric performance testing and reporting — Part 1: Principles and framework*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO/IEC 19795-1:2006 and the following apply.

3.1
influencing factors
factors that have influence on performance

3.2
robustness test
test to evaluate how much a certain influential factor affects biometric performance

3.3
active forgery attempt
impostor attempt in which an individual tries to match the stored template of a different individual by presenting a simulated or reproduced biometric sample, or by intentionally modifying his/her own biometric characteristics

3.4
forgery type
type of method of impostor forgery attempts

3.5
forgery level
difficulty level of impostor forgery attempts

4 Symbols and abbreviated terms

ROC receiver operating characteristic

CMR cumulative matching rate

CMC cumulative match characteristic

FNMR false non-match rate

FMR false match rate

5 Modality-specific test design

5.1 Development flow

When designing performance evaluation tests, the following test planning sequence should be used to ensure that the primary modality-dependent factors are accounted for (see Figure 1):

- Step 1: Identify and analyze "influencing factors" that may impact performance
- Step 2: Consider and establish policy on test subjects
- Step 3: Consider and establish policy on data collection
- Step 4: Consider and establish policy on impostor transactions
- Step 5: Consider and establish policy on performance reporting

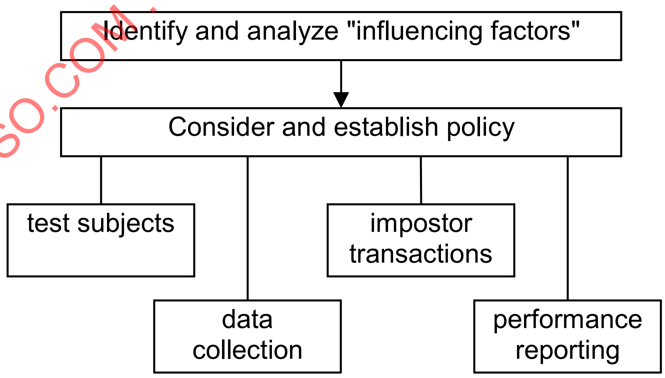

Figure 1 — Development flow

This sequence assumes that the modality and test type (i.e. technology / scenario / operational) have been defined.

5.2 Modality-specific factors that may impact performance (Step 1)

Factors that influence performance ("influencing factors") should be identified and analyzed in Step 1 of the test design development flow because biometric performance can be greatly affected by a wide variety of influencing factors. The same biometric device may generate different test results if these influencing factors differ. Controlling, recording, and reporting influencing factors is indispensable to executing repeatable performance tests and to predicting operational performance.

When developing test procedures, the test designer should identify and analyze those influencing factors known to impact performance for a given modality. The designer may also include modality-specific factors whose impact on performance is postulated but not known.

In identifying influencing factors, the designer should consider, at a minimum, the following (see also Annex C of ISO/IEC 19795-1:2006:

A) Biometric sensor quality and characteristics
B) Biological or behavioural characteristics of the subject relevant to data collection (essential historical or demographic data):
- Invariable: Gender, ethnic origin, occupation
- Variable:
  - ◆ Biological: age, body dimensions / anthropometric data (height, weight, etc…), musculoskeletal disorders
  - ◆ Habitual / Social factors: smoking preference, hairstyle, makeup, eyewear (glasses, contacts, etc…), clothing

C) Environmental factors applicable to the biometric device, sensor, or application such as:
- Temperature
- Humidity
- Illumination
  - ◆ Type (standard incandescent, fluorescent, tungsten halogen, reflector lamps, light emitting diodes (LEDs), sunlight, etc…)
- Noise
- Position of sensor with regards to the user

D) Temporal change of the biometric features
E) Impact of active forgery attempts on false acceptance, particularly in behavioural modalities
F) Differences between the data capture and signal processing subsystems used in the enrolment phase and those used in the verification/identification phase

Clause 6 of this document describes "robustness tests", modality-specific tests that evaluate the effects of environmental factors described in (C) above.

5.3 Modality-specific policies pertaining to Test Subjects (Step 2)

5.3.1 Policies pertaining to Test Subjects

Policies and requirements related to the Test Subject should be considered and established in Step 2 of the test design development flow. These policies will be in the area of biological and behavioural characteristics. Subject-specific influencing factors that should be considered include but are not limited to the following:
- Historical or demographic data
- Body dimensions / anthropometric data

The following sub-sections describe the distribution of test subjects that each evaluation should consist of. Table 1 shows the general relationship between crew composition and evaluation type. The following subsections will provide specific details that should be followed for each evaluation type.

To ensure an evaluation of the subject-specific influencing factor(s) consists of a distribution similar to the population required (general population, target users, etc…) standard anthropometric tables such as in [10] should be utilized where it is appropriate for subject-specific influencing dimensions. For variable historical or demographic data census or other survey data should be used.

One way to maintain a representative distribution for a general population when conducting technology or operational tests is to use the following two values for body dimensions: 5% female for the lower limit and 95 % male for the upper limit (provided the influencing dimension is lower in females than in males). These two values are reported to accommodate most males and females in a "general population".

The distribution for scenario testing is more complex as the test simulates a real world application which the test subjects should represent the target users.

To ensure the subject-specific influencing factor(s) are representative of the target or intended application users, subject-specific influencing dimensions should be known and provided to the testing organization. If dimensions are not available, the testing organization should request these ranges for the targeted users in the simulated environment under evaluation. If this information is not provided or unavailable, anthropometric tables such as [10] provide measurement estimates for various ethnic populations, ages, and occupations and should be used. If the subject-specific influencing factors are historical or demographic data the distribution should be similar to that of the real environment

Table 1 — Relationship of crew composition and evaluation type

Evaluation Type	Crew Composition
Technology	Target population
Scenario	Target users for a specific application.
Operational	No control / limited control of users (general population)

5.3.2 Technology evaluation

In technology evaluations, the composition of the test corpus data utilized should mirror that of the target population (which in some cases is the general population) to ensure representative occurrence of test subject-specific influencing factors that impact performance. The test designer should establish policies and requirements as follows:

1. The test subjects should include individuals that exhibit the subject-specific influencing factors of the general population.

2. All data exhibiting the test subject specific influencing factors should be recorded and reported.

EXAMPLE: The height of individuals in a standing face recognition application.

1. The distribution of heights of all test subjects should be similar to that of the general population (for example, 5 % female to 95 % male).
2. Height measurement data should be recorded and reported to ensure the test is using a general population.

5.3.3 Scenario evaluation

In scenario evaluations, crew composition should mirror that of the target application user group to ensure representative occurrence of test subject-specific influencing factors that impact performance. The test designer should establish policies and requirements as follows:

1. The test subjects should include individuals that exhibit the subject-specific influencing factors of the target application.
2. All data exhibiting the test subject specific influencing factors should be recorded and reported.

EXAMPLE: The height of individuals in a standing face recognition application.

1. The distribution of heights of all test subjects should be similar to the target users found in the real-world application.

2. Height measurement data should be recorded and reported to ensure the test is following a similar distribution to that of the users of the real-world application.

5.3.4 Operational evaluation

In operational evaluations, crew composition may not be under the control of the test designer because the crew is comprised of actual users. The experimental setup of the evaluation area should be designed such that the design accommodates a general population (for example, 5% female and 95% male), unless otherwise specified by the client or specifications from the device manufacturer. Recording and reporting the distribution of test subject-specific influencing factors is nevertheless important to ensure that results are repeatable and fully documented. The test designer should establish policies and requirements as follows:

1. The test subjects should include individuals that exhibit the subject-specific influencing factors of the general population.
2. All data exhibiting the test subject specific influencing factors should be recorded and reported.

EXAMPLE: A test is conducted in an operational environment whose users are overwhelmingly female, with gender being an influencing factor.

1. The distribution of gender cannot be controlled due to the nature of the evaluation.
2. Gender data should be recorded and reported for repeatability, generalizability, and validity of the test such that readers understand to what degree the crew was representative.

5.4 Modality-specific policies pertaining to data collection (Step 3)

5.4.1 Policies pertaining to data collection

Data collection requirements should be considered and established during Step 3 of the test design development flow. Influencing factors that should be considered include but are not limited to the following:

- biometric sensor quality and characteristics
- environmental factors surrounding the biometric device
- temporal change of the biometric features
- impact of the human-biometric sensor interaction
- impact of active forgery on false acceptance (see the next clause)

Policies vary according to test types as follows.

5.4.2 Technology and scenario evaluation

In technology and scenario evaluations, data-collection-specific influencing factors are often controllable. The designer should thus establish policies for data collection-specific influencing factors. The designer should further require recording and reporting of data-collection-specific influencing factors if the test operator is unable to conduct data collection in the specified fashion.

5.4.3 Operational evaluation

In operational evaluations, data-collection-specific influencing factors cannot be controlled because the device is operated by actual users. Recording and reporting information related to data-collection-specific influencing factors is nevertheless important to ensure that results are repeatable and fully documented. The test designer should therefore require that data-collection-specific influencing factors be recorded and reported.

5.4.4 Multi-instance policies

5.4.4.1 General

For certain modalities, multiple instances can be collected from each test subject. For example, 10 fingerprints, 2 irises, more than 2 vein patterns and 2 hand-geometry instances can be collected from one test subject. In context-dependent behavioural modalities such as text-dependent voice verification, many "voiceprints" can be collected from each test subject.

For certain modalities, such as fingerprints, multiple instances from different body areas of a given test subject are known to have correlations. For such modalities that could have correlations, policies on whether different instances can be used for genuine and/or impostor attempts should be defined.

5.4.4.2 Examples: fingerprint

Policies may be established as follows, where (A-1, B-2) means (“Subject A’s Finger 1” compared against “Subject B’s Finger 2”) and A-1-2 means the 2nd fingerprint sample of Subject A’s Finger 1.

- Example 1. In genuine tests to establish FNMR, use of 10 fingers from the same test subject as independent samples may be permitted, such that (A-1-1, A-1-2) and (A-2-1, A-2-2) are valid cases.

- Example 2. In impostor tests to establish FMR, use of different fingers from the same test subject (e.g. A-1, A-2) is generally not permitted, as subject A’s finger 1 and finger 2 cannot be considered independent.

- Example 3. In impostor tests to establish FMR, (A-1, B-1) and (A-1, B-2) may be permitted. Allowing such usage could dramatically reduce the cost and effort of testing, particularly when testing systems with low observed FMR.

5.4.4.3 Examples: voice and signature

In test design for behaviour-based modalities such as text-dependent voice and signature, it is necessary to consider how the text is used in authentication and whether the impostor knows the text. Data collection and reporting policies may vary in accordance with these variations as the following examples:

Example 1: Technology evaluation of text-dependent voice authentication

Utterance should be based on the same text as the enrolled reference both in genuine tests and in impostor tests. The samples containing different utterance should not be used in the evaluation. It should be recorded and reported whether the impostor trials include cases where the enrolled data by a male speaker is tested by a female impostor and vice versa.

Example 2: Technology evaluation of text-independent voice authentication

Utterance can differ from the enrolled reference both in genuine tests and in impostor tests. It should be recorded and reported whether the impostor trials include cases where the enrolled data by a male speaker is tested by a female impostor and vice versa.

Example 3: Technology evaluation of text-prompted voice authentication

Utterance should be based on the same text as prompted by the system both in genuine tests and in impostor tests. How the prompted text is generated and controlled should be recorded and reported as much as possible. It should be recorded and reported also whether the impostor trials include cases where the enrolled data by a male speaker is tested by a female impostor and vice versa.

5.5 Modality-specific policies in impostor transactions (Step 4)

5.5.1 Policies in impostor transactions

To ensure accurate evaluation of behavioural modalities such as voice or signature, it is necessary to consider impostor effort, as results may be influenced by the degree of effort with which an impostor tries to imitate another test subject's behaviour. For example, the operational performance of behavioural modality-based authentication is known to be impacted by the type and degree of impostor efforts. A minimum of four types of impostor forgery attempts can de described, as shown in Table 2.

Table 2 — Levels of impostor forgery attempt

Forgery Type	Description
Random forgery (zero-effort impostor attempt)	Forger submits his/her own biometric characteristics as if he/she were attempting successful verification against his/her own template.
Simple forgery	Forger has knowledge of what to submit as biometric characteristics, but makes no attempt to simulate the biometric characteristics, for example, tracing a genuine signature in the case of signature-based authentication.
Simulated forgery	Forger copies the original biometric characteristics.
Skilled forgery	Forger imitates the static and dynamic information of biometric characteristics, often with observation and practice.

It is reasonable to suspect that random-forgery FAR will be low relative to that of other forgery types. In the case of voice verification, performance may be impacted if an impostor hears the authorized user's voice prior to the impostor attempt. Accuracy may be further impacted if the impostor practices imitating the authorized user's voice. The likelihood of false acceptance may increase further if the impostor is a relation such as a same-gender sibling, a twin, a parent, or a child. In the case of gait authentication, the result may be different if an impostor has a chance to see the authorized user's manner of walking or an impostor practices imitating the authorized user's walking.

It is thus necessary to take the degree of impostor efforts into consideration to obtain reliable testing results in behavioural modalities. Impostor data collection policies should be decided during planning stages. Such data collection decisions and policies must be described and reported.

In technology testing and scenario testing, random forgery should not be the basis of impostor attempts. In operational testing, random forgery attempts may be permissible if this is consistent with the manner in which impostors would attempt to mimic the enrolled data. Because test results depend heavily on the type and degree of forgery, the report should detail the type of forgery data used in the operational test.

5.5.2 Example: signature

One way to take the degree of impostor efforts into consideration in testing of signature-based authentication is to use the categorization of forgery types as in Table 3.

Table 3 — Example of signature forgery level categorization

Level	Description	Types
0	zero effort, forger signs a random name	"blind" forgeries (forger does not have access to copy of signature)
1	forger has heard name but not seen it in print (does not know spelling Steven vs. Stephen or Jon vs. John)	
2	forger has seen name in print such as a phone directory or business card but has not seen signature	
3	single sample, forger has access to a single sample of victim's signature (receipt or check)	static forgeries (forger only has access to signature after it is signed, does not observe signature creation)
4	multiple sample, forger has access to multiple samples of victim's signature (possibly additional sample of writing such as a hand written note)	
5	single observation, forger has observed the victim signing his or her name once	observation based forgeries (forger actually observes signature creation)
6	multiple observation, forger has observed the victim signing his or her name multiple times (possibly video tape of signing replayed over and over)	
7	"victim assisted" forgery, victim intentionally coaches forger to dynamically imitate signature	assisted forgeries (either victim or technology assisted forgeries)
8	"technology assisted" forgery, forger has access to digitizer output of victim's signature and is allowed multiple practice attempts to imitate speed, pressure and curves	

In signature-based authentication, performance testing should be based on skilled impostor attempts. However, in practice it is difficult to collect skilled impostor attempts from a large number of impostors for each test subject. A practical solution is to collect static and dynamic signature data (i.e. both the written word and the dynamic characteristics of handwriting, such as pen-trajectory, pen-inclination and pen-pressure) from each test subject, using the same word or words for all subjects. A pair of samples collected under this policy can be used for genuine attempts if the pair is from the same test subject or for impostor attempts if the pair is from two different test subjects. The impostor attempt can be classified either as simple, simulated or skilled forgery depending on the degree of subject's training beforehand. Multiple signature designs should be used, since performance is greatly influenced by signature shape or designs.

Example 1: Technology evaluation of signature authentication

Impostors often have knowledge of the content of the genuine signature (what is written in the signature) to some extent, and the forgery level of the impostors should be recorded and reported. Examples of forgery level categorization are in Table 2 and Table 3.

Example 2: Operational evaluation of signature authentication

Impostors should have knowledge of the full-name of the user, since signature is based on the name in most of the cases.

5.5.3 Example: voice

In testing text-dependent and text-prompted voice verification, similar considerations apply. However, since the effect of non-random forgery is not so obvious in speech-based authentication, use of mere simple forgery is sufficient to constitute impostor attempts.

Example 1: Operational evaluation of voice authentication

The speakers, both the genuine and impostors, can make any utterance. In genuine tests, the cases should not be excluded from the test where the user forgets the text or erroneously makes utterances.

5.6 Modality-specific reporting policies (Step 5)

For modalities whose primary use is in identification systems (e.g. surveillance and AFIS), or whose application often requires operator validation of results, cumulative matching rate (CMR) is an important measure for evaluating system performance. CMR can be defined as the probability that an identification system will successfully rank two sample features from the same individual within a given similarity range. CMR is in effect an index indicating the probability that the correct choice is placed within a certain rank in the identification result.

Such performance can be represented in a cumulative match characteristic (CMC) graph showing changes in CMR by rank. As shown in Figure 2, CMR representing the probability of the person of interest being identified within a given rank is plotted on the vertical axis, against rank on the horizontal axis. In addition, the accuracy limit as calculated based on the size of the database used in the test is rendered to indicate the reliability of evaluation results. If CMR exceeds the accuracy limit, it does not have sufficient statistical reliability.

The stability of an identification system with regard to influencing factors can be expressed through CMC variations obtained from robustness tests. CMC variations represent how CMR varies in conjunction with influencing factors. CMC variations can be demonstrated by a graph as shown in Figure 3, in which classes corresponding to parameter changes within a given influencing factor are plotted on the horizontal axis against CMR on the vertical axis.

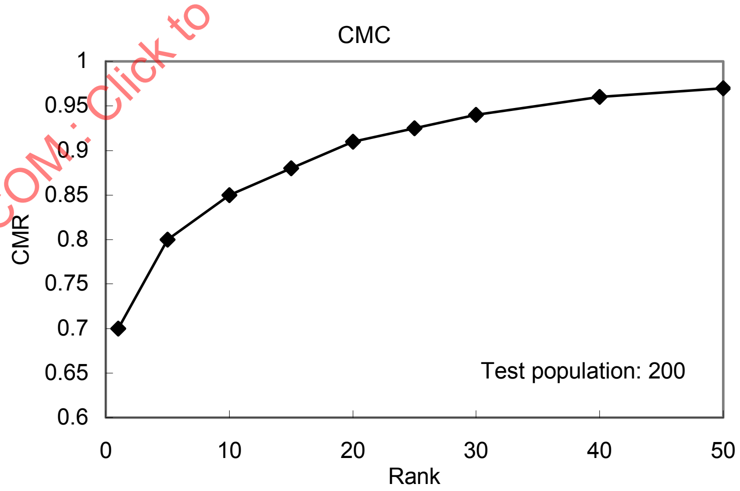

Figure 2 — CMC graph

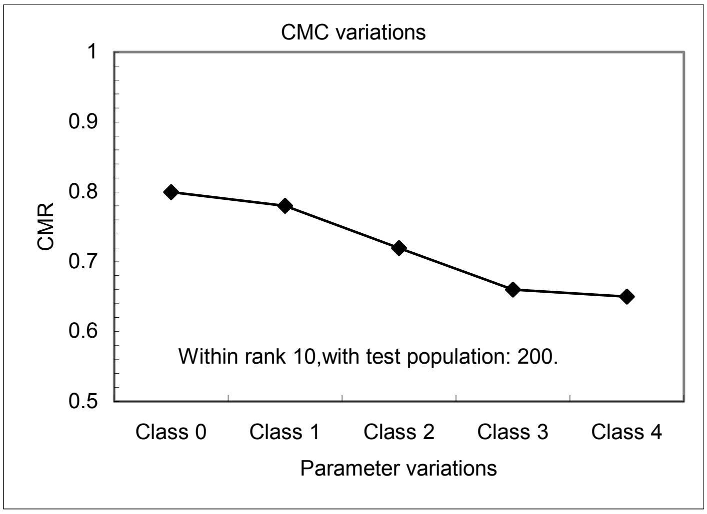

Figure 3 — CMC variations

6 Evaluation of modality-specific influential factors

6.1 Robustness tests

This clause describes “robustness tests”, which is an optional, modality-specific test to evaluate the effects of influential factors, for example, biological, social and environmental, that are analyzed in Step 1 of the test design process.

In general, technology testing does not consider the effects of environmental factors. However, some modalities, such as face and voice, can be greatly affected by a wide variety of factors. To quantify these impacts on performance, a “robustness test” can be introduced, which is designed to clarify and quantify the effects of each influential factor that can impact performance. The test is supposed to be used in parallel with technology tests and scenario tests.

A robustness test can be used to find out what kind of factors influence the performance and by how much, in other words, how sensitive the performance is against the change of each influential factor.

For example, in face authentication systems, the authentication performance could be affected by a wide variety of factors, including:

A) Biological or behavioural characteristics of the subject:
- Invariable factors: Gender, ethnic origin, occupation
- Variable factors:
  - ◆ Biological: age, body dimensions / anthropometric data (height, weight, etc…), musculoskeletal disorders
  - ◆ Habitual / Social factors: smoking preference, hairstyle, makeup, eyewear (glasses, contacts, etc…), clothing

B) Environmental factors applicable to the biometric device, sensor, or application such as:
- Illumination
  - ◆ Type (standard incandescent, fluorescent, tungsten halogen, reflector lamps, light emitting diodes (LEDs), sunlight, etc…)
  - ◆ Illumination source position variations (such as above, below, left, right and behind.)
- Position of sensor with regards to the user

In general, testing the influence on performance caused by such factors can be reduced by introducing specific “scenarios” in the test protocol. For example, if finger rotation degrades the fingerprint authentication performance in some systems, this kind of influence is usually minimized by requesting the users to place his or her fingers in an appropriate way.

However, making users follow these kinds of scenarios does not always work well in face authentication, because face-based systems are often expected to be used in applications in which subjects are not necessarily cooperative, such as surveillance. It has been observed that there is a strong demand for robustness tests from testers, i.e., system users who need to test the stability of the biometric system performance, by examining the degree of performance degradation resulting from changes in various factors.

The robustness of a biometric system can be reported by ROC curve variations on verification results. These ROC variations are demonstrated by a graph, (Figure 4) that represents how error rates such as FNMR and FMR change along with parameter variations.

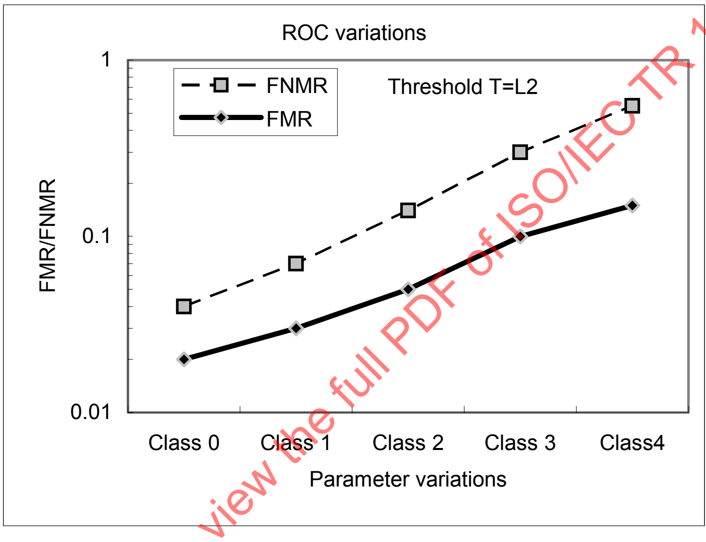

Figure 4 — ROC variations

6.2 Base example: face

Testing all environmental factors is not realistic, so the administrator should choose the relevant environmental factors and their specific ranges from the scenario before testing. Also, the administrator should describe the data acquisition conditions. Table 4 shows examples of some of the principal influential factors that relate to face-based authentication systems.

Table 4 — Example of principal influential factors (the case of face-based authentication systems)

	Parameters	Description of parameters	Influence on error rates	Cooperation of the test subject	Quantification	Requirements for standardization	
						Basic test	Robustness test
Biological factors	Genetic factors	Variations in the face geometry and colour due to genetic factors	Error rates increase if the features of samples differ from the designed/learned ones.	Impossible	Difficult	Indicate the place (country, city), date, and time of image capturing.	Follow the procedures for the basic test.
	Health status	Change in face geometry and colours caused by illness and injury	Error rates increase if the features and conditions of samples differ from the designed/learned ones.	Impossible	Difficult	No need for description	No need for description
	Age	Change in face geometry, skin elasticity, wrinkles and colours caused by aging	Error rates increase if the features of samples differ from the designed/learned ones.	Impossible	Possible	Indicate the distribution for each feature.	For comparison with the basic performance evaluation test (neglecting the influence of aging), indicate the matching performance specific to each distribution.
	Gender		Error rates increase if the features of samples differ from the designed/learned ones. Note that this factor is often highly correlated with other factors.	Impossible	Possible	Indicate the distribution of each feature (for men and women).	For comparison with the basic test (neglecting gender-specific differences), indicate the matching performance for each gender.
	Expression	Change in face geometry and wrinkles caused by a change in expression	Error rates increase if the features and conditions of samples differ from the designed/learned ones. The less expression, the greater the success rate.	Possible	Difficult	Indicate the presence/lack of cooperation and the relevant details.	1) Indicate the dependency of performance on the request for specific expression, or 2) on the expression categories the administrator has set.
Social factors	Occupation		Error rates increase when some specific changes occur to the appearance. This is, however, not as distinctive as changes in fingerprint matching.	Impossible	Difficult	Indicate the distribution of each feature.	Indicate the matching performance of each occupational group for comparison purposes. The administrator may set the occupation categories as desired.
	Hairstyle/ Beard/ Makeup, etc.	Including accessories such as beards, hairstyles covering facial parts, makeup on lips and eyebrows, tattoos, shadow, highlight, eye patches, surgical masks	Error rates increase if the features of samples differ from the enrolled features.	Possible	Difficult	Unavailable data may be removed. Indicate different conditions and availability.	The administrator determines the categories and indicates the performance variations between categories.
	Eyeglasses		Error rates increase if the features of samples differ from the enrolled features. Error rates are also affected by changes in eyeglasses and the resultant shadows and reflection.	Possible	Difficult	Indicate the ratio of persons wearing eyeglasses. Unavailable data may be removed, but availability must be recorded.	Indicate performance variations resulting from appearance changes due to eyeglasses being worn/not being worn; indicate model images.

	Parameters	Description of parameters	Influence on error rates	Cooperation of the test subject	Quantification	Requirements for standardization	
						Basic test	Robustness test
Environ-mental factors	Posture	Orientation of face in relation to the camera	Error rates increase if the features of samples differ from the enrolled features. A frontal view of the face is determined to be the normal condition. Off-angle or side views lead to increased error rates.	Possible	Possible	Substantially fix the posture in relation to a fixed camera during image capturing. Indicate the presence/lack of cooperation and the relevant details.	The administrator classifies the relative positions between the camera and the face, and indicates performance variations among the classes.
	Illumination	Direction of illumination and number of lights	Error rates increase if the features of samples differ from the enrolled features.	Impossible	Possible but requires coupling with other parameters	Substantially fix the direction. Determine the illumination conditions using a ball that has Lambertian reflection property (e.g. a plaster ball).	Test each illumination condition, and indicate performance variations between the plaster-ball conditions and actual illumination conditions.
	Background	Background during image capturing	A complex background making separation difficult causes error rates to increase.	Impossible	Possible but requires coupling with other parameters	Fix the background. Show the background of the sample clearly.	Indicate performance variations caused by background changes; indicate background images.
	Resolution	Distance and resolution during image capturing	Error rates increase if the features of samples differ from the enrolled features.	Impossible	Possible	Indicate the camera FOV, pixel number, and distance to the face.	Indicate performance variations among parameters.
	Time interval	Time from enrolment to authentication	Error rates increase if there is a prolonged time interval.	Impossible	Possible	Fix the time intervals. Record the shortest time interval between enrolment-data acquisition and verification-data acquisition.	Plot the matching performance in relation to the time from enrolment to matching as a parameter.
	Distortion	Distortion caused by properties of the camera	Error rates increase if distortion differs among the samples.	Impossible	Impossible	Indicate the camera type.	Indicate the performance variations among different cameras.

6.3 Other examples: fingerprint, iris, vein, and voice

When administrators carry out tests for other biometric modalities, the administrator shall consider effects of other factors. Taking fingerprint recognition; humidity, skin condition (Table 5), etc. shall be considered. For iris recognition, the illumination condition, eyewear and eye disease shall be considered (Table 6). For vein recognition, changes in vascular patterns, hand/finger orientation, and environmental illumination shall be considered (Table 7). Example of principal influential factors in case of voice authentication is in Table 8.

Table 5 — Example of principal influential factors (the case of fingerprint-based authentication systems)

	Parameters	Description of parameters	Influence on error rates	Cooperation of the test subject	Quantification	Requirements for standardization	
						Basic test	Robustness test
Biological factors	Health status	Change in finger surface (cut, slash, gash, scratch, etc.) caused by injury	Error rates increase if the features and conditions of samples differ from the designed/learned ones.	Impossible	Difficult	No need for description	No need for description
	Age	Change in finger shape, skin elasticity, wrinkles caused by aging	Error rates increase if the features of samples differ from the designed/learned ones. However, in general, young adults provide successful results.	Impossible	Possible	Indicate the distribution for each feature.	For comparison with the basic performance evaluation test (neglecting the influence of aging), indicate the matching performance specific to each distribution.
	Skin condition	Change in finger skin condition, wet or dry, caused by biological or physiological reasons.	Error rates increase if the features and conditions of samples differ from the designed/learned ones. The less the condition changes, the greater the success rate.	Possible	Difficult	Indicate the presence/lack of cooperation and the relevant details.	1) Indicate the dependency of performance on the request for specific skin condition, or 2) on the skin condition categories the administrator has set.
	Dominant hand or not	Motion of finger changes when using the sweep scanner.	Error rates increase if the subject is using the non-dominant hand	Possible	Possible	Indicate the use of dominant or non-dominant hand.	Same as basic test.
	kind of finger (size of finger, etc)	Which finger to use for authentication	Error rates increase when, for example, ring or little fingers are used.	Possible	Possible	Indicate which finger to be used.	Indicate the matching performance for different kind of fingers.
Social factors	Occupation and lifestyle	Change in finger skin condition, roughness, scratch, keratinized or chaps, due to occupation and hobbies (gardening, ceramic art, sports, holiday carpenter, etc.).	Error rates increase if the skin condition is worse.	Impossible	Difficult	Indicate the distribution of each feature.	Indicate the matching performance of each occupational group for comparison purposes. The administrator may set the occupation categories as desired.
Environ-mental factors	Time interval	Time from enrolment to authentication	Error rates increase if there is a prolonged time interval.	Impossible	Possible	Fix the time intervals. Record the shortest time interval between enrolment-data acquisition and verification-data acquisition.	Plot the matching performance in relation to the time from enrolment to matching as a parameter.
	Distortion	Distortion caused by properties of the scanner	Error rates increase if distortion differs among the samples.	Impossible	Impossible	Indicate the scanner type.	Indicate the performance variations among different scanners.
	Scanner	Scanner properties (except for illumination, background, resolution, and distortion), time, and number of scans per attempt	Error rates increase if the features of samples differ from the enrolled features.	Impossible	Difficult	Fix the conditions. Indicate the presence/lack of functions such as auto white-balancing.	Indicate the performance variations among different fingerprint scanner systems.
	Temperature and Humidity	Temperature and Humidity that affect skin condition and the scanning performance.	Error rates changes if the skin condition or the scanning performance is affected by the parameters.	Possible (partly, for example, dry finger by wiping with cloth)	Possible	Report the parameters.	Report the performance changes due to parameter variations.