B. Error rates

The interaction with a biometric system starts with the enrolment, where the quality of enrolment data is very important and significantly influences the system performance. Often several input samples (e.g. 3 or 5) are combined to create one biometric reference (or to verify usability of the newly created biometric reference).

The probability of a person not being able to enrol in a biometric system is called the Fail to Enrol rate (FTE). It is computed as a fraction of people who could not enrol in the system out of the complete group of people. The FTE rate includes people without fingers (for fingerprint systems), visually impaired people (for iris-based systems), etc.

For verification/identification attempts, the biometric input sample is obtained and its quality is verified. If the quality does not satisfy certain minimal quality requirements, the acquisition process must be repeated. If all repeated acquisitions do not yield sufficiently good samples, the person cannot be identified/verified and such an attempt increases the Failure To Acquire (FTA) rate. Sometimes the minimal quality can be configured and then it is clear that the stricter we are with the quality check the better result we get during the biometric comparison and vice versa. The FTA rate can be therefore traded off with biometric matching error rates.

Input samples of sufficient quality are processed in the biometric matching algorithm. The matching algorithm compares the input sample with a biometric reference (in the case of verification) or number of references (in the case of identification). The result of the matching algorithm is either correct or incorrect. If an error occurs, the resulting decision can either incorrectly refuse an authentic person (this is so-called false non-match – FNM) or match an impostor with another person's biometric reference (this is so called false match – FM). What happens next depends on the system policy. In the case of single attempt scenario, the verification/identification ends. In the case of, for example three-attempt scenario, re-acquisition is possible if the person is not being recognized (either false non-match or correct refusal of an impostor).

The final result of an authentication/verification attempt is either correct acceptance or correct refusal, false acceptance or false rejection. In the case of single-attempt scenario the FRR and FAR can be computed as:

$$FRR = FTE + (1 - FTE) \cdot FTA + (1 - FTE) \cdot (1 - FTA) \cdot FNMR$$

$$FAR = (1 - FTE)^2 \cdot (1 - FTA) \cdot FMR$$

For the purpose of FAR computations the so-called *zero-effort* (also called *random forgery*) unauthorized authentication attempts are taken. In this case attackers are not actively changing their biometric characteristics (for example in the case of dynamic signature systems they sign as usual).

Sometimes the minimal quality required for a successful enrolment can be configured. It is however clear that the stricter we are with the quality control at the time of enrolment (i.e. the better quality of the biometric reference), the better results we achieve later in verification/identification attempts and vice versa. Therefore matching error rates can be traded off with the enrolment quality requirements. In 2004 Atos Origin (commissioned by the UK Passport Service) ran a biometric trial. Facial, iris and fingerprint systems were tested in real conditions with 3 groups of participants: Quota (representative sample of the population), Opportunistic (volunteers) and Disabled (several types of disabilities). The Quota and Disabled results can be briefly summed in the table 1. For details (explanation of some of the results, shortcomings of the trial etc.) see the final report of the trial [5].

	Face				Iris				Fingerprint			
	FTE	FTA	FNMR	FRR	FTE	FTA	FNMR	FRR	FTE	FTA	FNMR	FRR
Quota	0.15	0.00	30.82	**30.92**	12.30	0.44	1.75	**14.21**	0.69	6.98	11.70	**18.43**
Disabled	2.27	0.00	51.57	**52.67**	39.00	0.68	8.22	**44.39**	3.91	3.14	16.35	**22.14**

TABLE I

THE ERROR RATES OF FACIAL, IRIS AND FINGERPRINT SYSTEMS IN A UK 2004 TRIAL [5].

ALL VALUES ARE EXPRESSED IN %.

The correct way to calculate error rates is to compute error rates for each person who contributes to the tests and then to average[3] the rates over the group of all the people. Otherwise the results can be biased by an unbalanced number of verification/identification attempts done by different people.

[3] Weighted average corresponding to the target population can also be used.

As we have seen, the accuracy/usability of biometric systems can be measured in the terms of FTE, FTA, FMR, FNMR, FAR and FRR. When comparing different systems, typically only the resulting FR and FA rates are used. The FAR and FRR can be graphically expressed in a FAR-FRR graph, where both the error rates are a function of the threshold value or can be plotted in a ROC graph where the FAR is a function of FRR or vice-versa (thus eliminating the threshold value from the graph). Figures 1 and 2 give a simplified example of such graphs. The point where FAR and FRR have the same value is called the equal error rate (EER) or the crossover accuracy. Such a threshold does not have a particular importance, but the resulting EER can be used as a (rather simplified) performance value of a biometric system in evaluations.

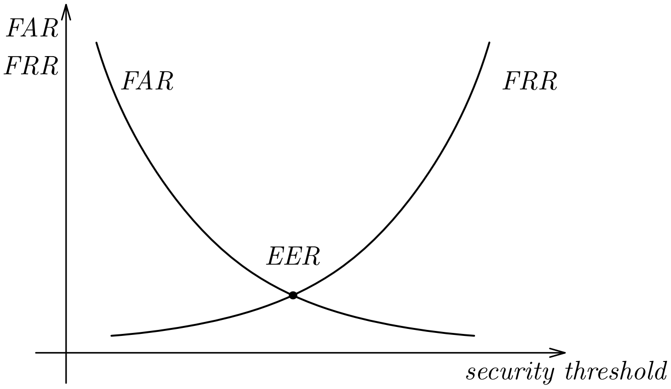

Figure 1: FAR-FRR graph (idealized).

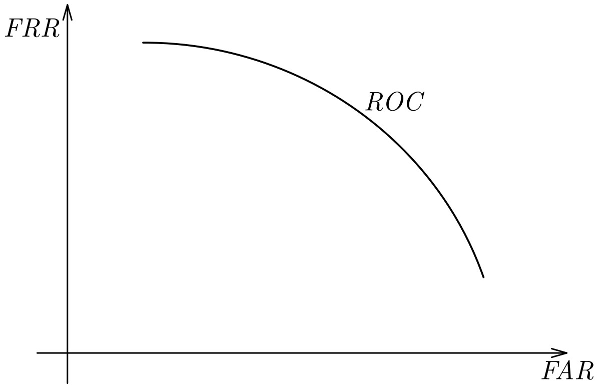

Figure 2: ROC graph (idealized).

Now let us review some real numbers. There are several types of tests [7] and not all the results must necessarily be comparable.

The American NIST is regularly testing the accuracy of fingerprint and facial biometric systems. As an example of the result of their test effort we include here the ROC graph of facial biometric systems from 2006. The details of the NIST tests can be found at `fingerprint.nist.gov` and `face.nist.gov`.

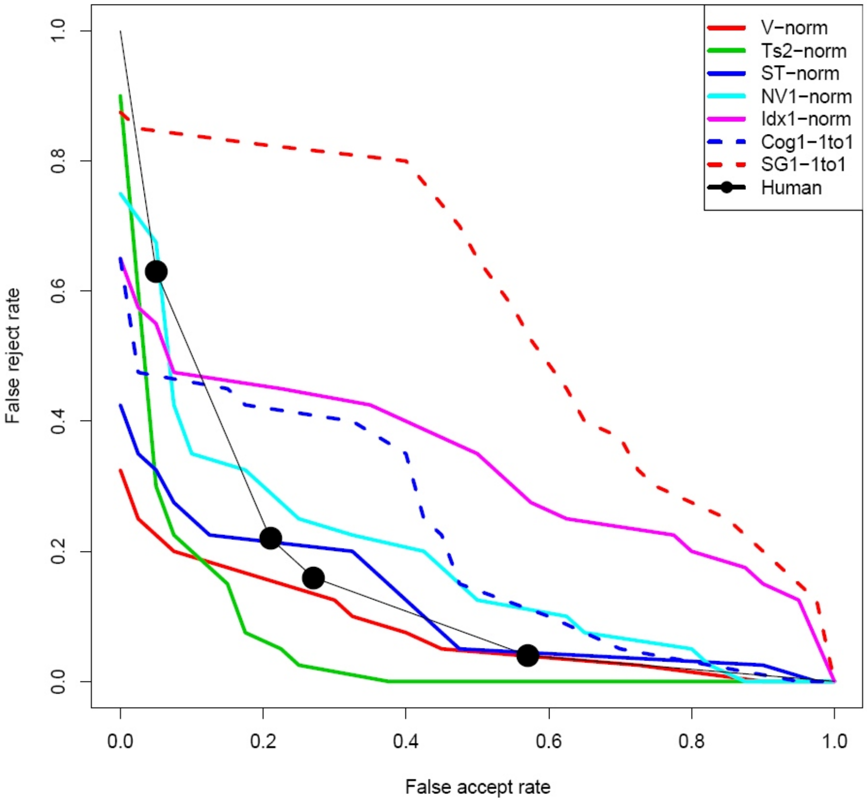

Figure 3: The ROC graph of several facial recognition algorithms and human ability to recognise faces (FRVT 2006 run by NIST [35] for facial images with illumination changes).

C. Large scale systems

Designing a biometric system for a few of data subjects (as users are called according to [23]) is relatively easy. Tuning a system for millions of data subjects is significantly more challenging.

While the *verification* speed and accuracy is essentially same for a system with 10 data subjects and for a system with 10 million data subjects, the *identification* mode makes the difference.

In identification mode the biometric system can incorrectly reject the data subject (and this affects the false-negative identification-error rate – FNIR) or incorrectly accept an impostor (and this is measured by the false-positive identification-error rate – FPIR).

In the case of a single attempt scenario the values of FNIR and FPIR can be estimated from