Biometrics Person Authentication: From Theory to Practice

Florent Perronnin

Institut Eurécom, 2229 Route des Crêtes, 193 - 06904 Sophia Antipolis Cedex, France

Xerox Research Centre Europe, 38240 Meylan, France
Email: florent.perronnin@xrce.xerox.com

Jean-Claude Junqua

Panasonic Speech Technology Laboratory, Santa Barbara, CA 93105, USA

AV Core Technology Development Center (ACC), Matsushita Electric Industrial Co., Ltd.,
Osaka 571 - 8506, Japan
Email: junqua.jean-claude@jp.panasonic.com

Jean-Luc Dugelay

Institut Eurécom, 2229 Route des Crêtes, 193 - 06904 Sophia Antipolis Cedex, France
Email: jean-luc.dugelay@eurecom.fr

Biometrics, which refers to identifying an individual based on his/her physical or behavior characteristics, is an emerging topic in the field of signal processing. Nevertheless, biometric applications have failed to become as widespread as once anticipated. After a very brief review of the field of biometrics, the present tutorial focuses on three main aspects of biometrics: maturity of technologies and algorithms, applications, and evaluations.

1. Introduction

There exists a wealth of applications that require reliable person identification or identity verification. The two traditional approaches to automatic person recognition, namely the *knowledge-based* approach which relies on something that one knows such as a password, and the *token-based* approach which relies on something that one has such as a badge, have obvious shortcomings: passwords might be forgotten or guessed and badges might be lost or stolen.

Biometrics, which is the discipline concerned with "the automatic identification or identity verification of an individual based on *physiological* and *behavioral characteristics*" [1, 2, 3, 4], is an alternative to these traditional approaches as a biometric attribute is inherent to each person and thus cannot be forgotten or lost and might be difficult to forge. Face [5, 6, 7], fingerprint [8, 9], hand/finger geometry [10], retina and iris [11], ear [12, 13] are examples of physiological characteristics while signature [14], voice [15, 16, 17, 18], gait or keystroke are examples of behavioral characteristics.

While biometrics has been the subject of decades of research, biometric applications have failed to become as widespread as once anticipated. This is mainly due to the following reasons.

- The first reason is the fact that biometrics is a very challenging pattern recognition discipline and that even state-of-the-art systems may not be mature enough, as shown in recent evaluations of fingerprint [19], face [20], and voice [21] technologies.
- The second reason is that most of the research effort on biometrics has focused on the technology aspect but that, in comparison, little thought has been given to the careful development of applications.
- The third reason is that evaluation of biometric systems has long been an overlooked issue. Indeed, evaluations generally take into account only the technology while they should also consider the application. Therefore, it is often difficult to predict the performance of a biometric system in the real world.

After a very brief review of the field of biometrics, we will consider the three challenges that have been previously outlined. We will then discuss the use of multimodality to alleviate some of the shortcomings of individual biometric systems. We will also consider an example of a successful application deployed in an R&D laboratory.

2. A brief review of biometrics

2.1. History

With the abolition of prisoner marking (e.g., 1832 in France), the identification of habitual offenders became harder. In 1880, Alphonse Bertillon (chief of criminal identification for the Paris police) proposed a system to identify persons by means of a detailed record of body measurements (arm, head, ear, etc.), physical description (color of eyes, scars, etc.), and photographs. Bertillonage was officially adopted in France in 1882 and soon after in some other countries. The National Bureau of Identification, forerunner of the FBI (Federal Bureau of Investigation), was established in Chicago in 1897, and owned 24 000 Bertillon cards after two years. The Bertillon system was generally accepted for thirty years, but the system was in fact so cumbersome that two different individuals measuring the same person frequently would not arrive at the same description; it was also difficult to administer the system in a uniform way. Finally, the case of Will West (i.e., two different persons having the same Bertillon measurements) strengthened the case in favor of the science of fingerprints as the normally accepted method of personal identification (in 1903, the NY State Prison system began the first systematic use of fingerprints in the US for criminals). The pioneer works and contributions in fingerprints are by Dr. Henry Faulds (1880) and Sir Francis Galton (1888). Even if fingerprints are efficient in a criminal environment (the odds are 67 billion to one against any two different persons producing identical prints), this biometric includes some weaknesses that make it inappropriate for some other applications (e.g., difficulties for some people to register, medium user acceptance, conditions of acquisition, etc.).

2.2. Properties

Ideally a biometric should have the following properties [22, 23].

- *Universal*: all the persons should have the characteristic.
- *Permanent*: the characteristic should not vary over time.
- *Distinctive*: samples corresponding to different persons should be as different as possible, that is, the interclass variability should be as large as possible.
- *Robust*: samples corresponding to the same person should be as close as possible, that is, the intraclass variability should be as small as possible.
- *Accessible*: the sample should be easy to present to the sensor.
- *Acceptable*: it should be perceived as nonintrusive by the user.
- *Hard to circumvent*: it should be hard for an impostor to fool the system.

2.3. Operational mode

It is of utmost importance to distinguish between the two main operational modes of a biometric system.

• In the *identification* mode, the user makes no claim of identity and the system has to perform a search over the entire database to find the most likely identity (one-to-many comparisons). A *close-set* is generally assumed which means that all the trials are supposed to be from persons who are registered in the database.

• In the *verification* mode, the user claims an identity and the system has to decide whether the sample indeed corresponds to the claimed identity (one-to-one comparison). An *open-set* is generally assumed, which means that the input samples may correspond to persons who are not registered in the database.

In the following sections, we use the generic term *recognition* when we do not want to make the distinction between identification and verification.

2.4. Architecture

Biometric applications involve typical pattern classification systems. The architecture of a generic biometric system is shown in Figure 1. It is composed of at least two mandatory modules, the *enrollment* and the *recognition* modules, and an optional one, the *adaptation* module.

• Enrollment is performed when a person registers in a biometric system. The typical stages of the enrollment are as follows. A *sensing* device first captures the biometric of interest. A series of *preprocessing* steps is then applied to the obtained signal. For the problem of face recognition, such preprocessing operations may include face detection/segmentation, geometric or photometric normalization, and so forth. A very important component of many preprocessors is the *quality checker*: if the quality of the input signal is too poor, the system may require another sample from the user. Then features are extracted from the signal. The goal of the *feature extraction* step is to extract the unique features that characterize the considered person while discarding irrelevant information. Finally, a *model can be estimated* with the available features. It is subsequently stored, for instance, on a smart card or in a centralized database.

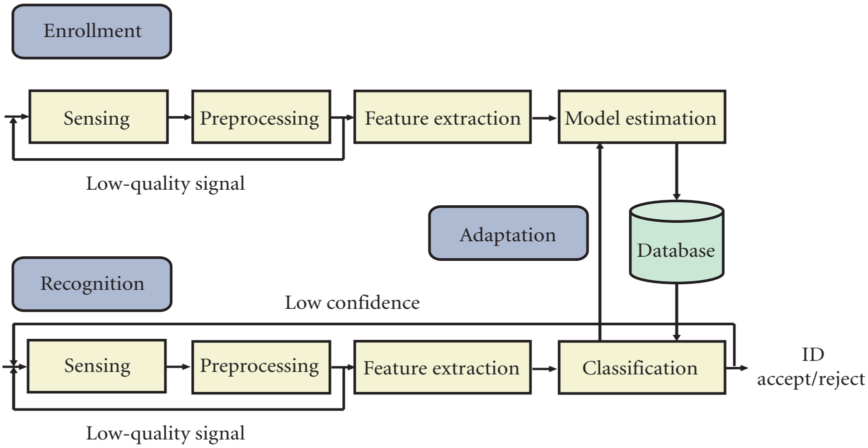


Figure 1: Architecture of a typical biometric system.

- The first steps of the recognition are generally similar to the ones of the enrollment: sensing, preprocessing, and feature extraction. Then one or multiple templates are retrieved from the database, depending on the operational mode. The extracted set of features is then compared with the template(s). Based on the outcome of the matching and the decision policy of the biometric system, a decision is taken. In the verification mode, the system can take an acceptance or rejection decision or, in a case of uncertainty, request additional data from the user.
- During the enrollment phase, a user-friendly system usually only captures a few instances of the biometric used. This is generally insufficient to describe with great accuracy the characteristics of this attribute. Moreover certain biometrics such as face and voice are not permanent. Hence, the goal of the adaptation module is to maintain or even improve the performance of the system over time by updating the model after each or several access to the system.

2.5. Performance measures

As the identification and verification are two different operational modes, they require different measures of performance.

- The *identification rate* is generally used to report the performance of a biometric system in the identification mode. If the top match corresponds to the identity of the person who submitted the query, then a success is declared. The identification rate is the percentage of such successful requests. Another measure of performance is the *cumulative match score*. A success is declared if the identity of the person who submitted the query is among the top N matches. The performance of a system can be represented graphically by drawing the cumulative match score as a function of N. When a search has to be performed over a very large database of templates, other performance measures can be considered such as the *penetration rate* and the *binning error rate*.

• When a biometric system works in the verification mode, it can make two types of errors. It can either reject a person that made a rightful identity claim, also referred to as a *client*, or accept a person that made a wrongful identity claim, also referred to as an *impostor*. The false rejection rate (FRR) is the expected proportion of transactions with truthful claims of identity that are incorrectly denied. The false acceptance rate (FAR) is the expected proportion of transactions with wrongful claims of identity that are incorrectly confirmed. Note that the FAR and FRR are defined over transactions. To avoid ambiguity with systems that allow multiple attempts or that have multiple templates per user, the false match rate (FMR) and the false nonmatch rate (FNMR) have been defined for a single comparison of a query against a single enrolled template. To take an acceptance/rejection decision, a biometric system typically compares the matching score to a decision threshold θ. If the matching score falls below θ, then the claim is considered wrongful. If the matching score is higher than θ, then the claim is considered rightful. Obviously, the FAR and FRR are conflicting types of errors. The system performance can be depicted in the form of a receiver operating characteristic (ROC) curve. It plots parametrically as a function of θ the FAR against the FRR. For a given application, θ should be set according to the desired level of security / convenience. The equal error rate (EER), which corresponds to the point where FAR = FRR, is often used to report the performance of a system. A decision cost function (DCF) may also be used to summarize the performance of a system with one unique figure for a given threshold θ: $\mathrm{DCF}(\theta) = C_{\mathrm{fr}}P_{\mathrm{clt}}P_{\mathrm{fr}}(\theta) + C_{\mathrm{fa}}P_{\mathrm{imp}}P_{\mathrm{fa}}(\theta)$. C_{fr} and C_{fa} are, respectively, the costs of a false rejection and of a false acceptance, P_{clt} and P_{imp} are, respectively, the prior probabilities of client and impostor attempts, and $P_{\mathrm{fr}}(\theta)$ and $P_{\mathrm{fa}}(\theta)$ are, respectively, the FRR and FAR for a given threshold θ.

3. Challenges

3.1. Technology issues

Biometrics is a very challenging pattern recognition task, both in the identification and in the verification modes. In the identification mode, depending on the application, the number of classes can be very large. In the verification mode, only two classes are considered, which correspond to the acceptance and rejection decision, but the challenge stems from the difficulty to properly model impostors who are generally unknown. Other challenges lie in the pattern recognition system itself, namely two modules: the *feature extraction* and the *classification* modules.

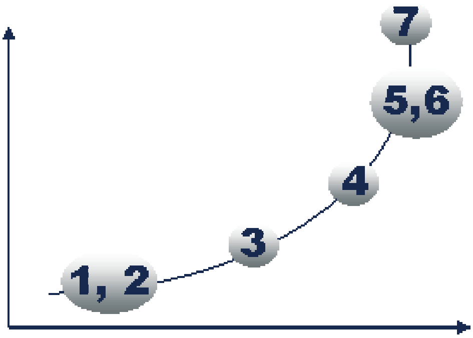
• For an efficient feature extraction, it is of paramount importance to understand precisely what characterizes the biometric. Toward this end, understanding how persons identify each other can be extremely beneficial, although this may not be sufficient. We take the example of automatic speaker verification (ASV). ASV has long focused on the extraction of low-level features, which provide information about the acoustic of speech. Such features are highly correlated with the *physical* traits of the user. Their great advantage is that they are easy to extract automatically. Their main downside is that they are not robust, especially to channel effects (which include different microphones, acoustic environments, or transmission channels). Most state-of-the art systems make use of *cepstral acoustic features* derived from the speech spectrum. They summarize *short-term* information as these features are typically extracted from 20 millisecond windows. More recently, it has been shown that *high-level* information, which characterizes the *behavioral traits* of the user, contains also a lot of valuable information for the problem of ASV. This includes, for example, choice of sequences of words and pronunciation [24, 25, 26]. Such high-level information is however more difficult to extract but seems to exhibit less sensitivity to channel or speaker variabilities.

• The goal of the classification module is to distinguish between intra- and interclass variabilities. This is a particularly challenging issue as biometric samples of different persons share global characteristics while biometric samples of the same person are subject to considerable variability, which might overwhelm the interperson differences. For instance, in the case of automatic face recognition (AFR), such variability is due to a long list of factors including facial expressions, illumination conditions, pose, presence or absence of eyeglasses and facial hair, occlusion, and aging. The classification problem is quite difficult as there is generally very little data to characterize the intrapersonal variabilities. Indeed, *data scarcity* is a problem of paramount importance in biometric applications, especially for behavioral characteristics that characterize a *dynamic*. When a new user first enrolls in a system, only a few instances of the considered biometric are typically captured in order to reduce the duration of enrollment and minimize inconvenience to the user (as well as maximize user cooperation). Hence, very little intraclass variability can be observed during the enrollment session. If only one sample is provided (e.g., one face image in the case of AFR), intraclass variability is obviously impossible to assess. To overcome this issue, one generally has to postulate that the intrapersonal variability is the same for all persons. Thus, the parameters of the model parameters can be estimated from a larger training set which is not restricted to the data of the person under consideration. However, this hypothesis is only a crude approximation.

3.2. Understanding the application

All applications are different and to successfully deploy a biometric system, it is necessary to fully understand the application requirements. Wayman suggests in [23] a partitioning of biometric systems according to the following seven categories.

- *Cooperative/noncooperative*. This terminology refers to the behavior of the deceptive user. In applications verifying a positive claim of identity, the deceptive user cooperates with the system in the attempt to be recognized. On the other hand, in systems verifying a negative claim of identity, the deceptive user will be noncooperative in the attempt not to be recognized.
- *Overt/covert*. A system is said to be overt if the user is aware that one of his biometrics is being measured. If not, the system is covert.
- *Habituated/nonhabituated*. This refers to the frequency of the interaction of a user with the biometric system.
- *Attended/nonattended*. This refers to whether the use of the biometric device is observed or guided by a person. Most systems supervise at least the enrollment process.
- *Standard/nonstandard environment*. This refers to the conditions of operation of a biometric system. For instance, outdoor systems will generally be considered as nonstandard environment applications.



Security level in function of the type authentication

1 Something you know (e.g., PIN code, birthday).
2 Something you have (e.g., card, key).
3 Something you know + something you have (e.g., ATM card + PIN).
4 Something you are.
5 Something you have + something you are.
6 Something you know + something you are.
7 Something you know + something you have + something you are.

Figure 2: The use of several sources of knowledge yield different security levels.

- *Public/private*. This refers to whether the users of the system will be customers or employees of the system management.
- *Open/closed*. This refers to whether the system will be required to exchange data with other biometric applications.

In every authentication system there is an apparent tradeoff between *user convenience* and *security*. A very secure system will have a higher rejection rate than a less secure one or alternatively it may have required several passes (such as the system described in Section 5) to increase security at the expense of user convenience. For some of the modalities (such as voice) the amount of enrollment data is directly related to this tradeoff as more enrollment data means generally a better tradeoff for the user. As shown in Figure 2, the security level depends of course on the type of biometric features used and the type of application but it also depends on the types of information used to authenticate a particular individual.

The tradeoff between security and convenience can also be illustrated by the tradeoff between false alarms and false rejections. Figure 3 shows, in the case of a text-dependent voiceprint identification, how the error rate varies when the system is tested (1) with the same password as the one uttered by the true user when the password is uttered by an imposter, (2) with a different password than the one uttered by the true user when the password is uttered by an imposter, and (3) with the same password as the one uttered by the true user when the password is uttered by the true user.

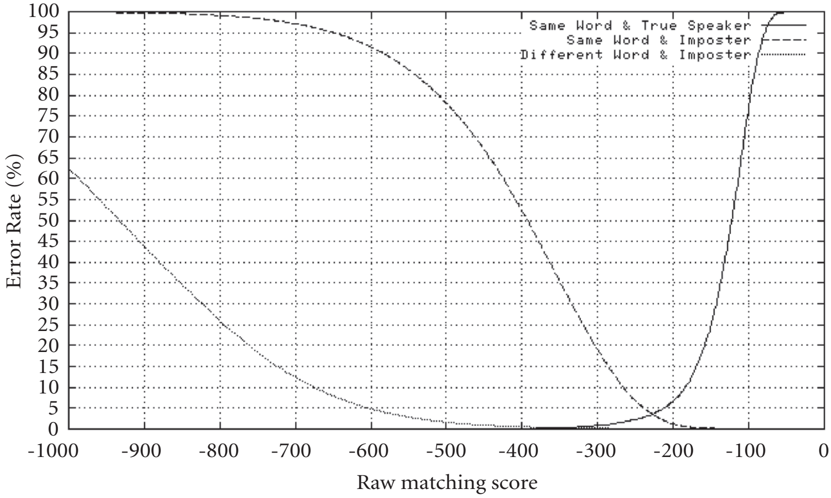


FIGURE 3: Error rate versus matching score for a text-dependent voiceprint identification system.

It can be seen that by choosing the right threshold on the raw matching score, it is possible to change the tradeoff between false alarms and false rejections. This has a direct incidence on the tradeoff between security and convenience. Each application requirements are different and consequently require a different tradeoff. To find the best tradeoff between security and convenience, it is necessary to adapt the security level to the local context and as a consequence the method to be used. *Understanding the application is very important and knowing in advance the type of users of the system can be one of the keys to the successful deployment of biometric systems [27].*

3.3. Performance evaluation

According to Mansfield and Wayman [27] biometric testing and evaluation can be of three types.

- Technology: by means of database evaluation.
- Scenario-based: in this case the overall system performance of a prototype or simulated application is evaluated. This allows in particular to test the human-machine interface.
- Operational: the performance of a complete system in a specific environment with a target population is evaluated. In general this type of evaluation is not repeatable.

Each type of evaluation has a separate purpose and produces different results. Unfortunately, the first type of evaluation is generally done and to some degree this limits the progress in this area. In particular the nature of impostors is an important part of biometric systems and generally the impostors used in technology evaluation are not true impostors. This explains why it is very difficult with today's technology to predict real-world error rates. In the example given in Section 5, the evaluation performed was an operational evaluation. This type of evaluation is costly but provides very good insight on the "true" performance of the system along with the areas to improve which are very often *not* directly related to the technology per say. Furthermore, in the evaluation performed in Section 5, manual examination of the audio files clearly indicates that intraspeaker variability (e.g., pitch, enunciation clarity, loudness, prosody changes) is not a negligible phenomenon. The natural variability is, however, difficult to measure. The phenomenon is exacerbated by the fact that users tend to pay the level of attention that is only needed to pass the authentication test. This highlights the fact that it is important to work on real data. Laboratory efforts often do not model behavior of target users in operational environment. Laboratory efforts mainly focus on improving the technology independently of the user behavior variability and the coupling between the user and the environment and the user and the task. This does not provide the entire picture and limits the progress that can be made. There is also a need to record data in a range of semicontrolled conditions that simulate real environments. Other factors to take into account besides equal error rates results include [28]

- failure and difficulties to enroll (e.g., amount of enrollment data needed) and failure to acquire across the test population (statistics show that 4% of fingerprints are of poor quality),
- reliability, availability, and maintainability,
- user acceptance and user convenience,
- human factors,
- vulnerability and ease of counterfeiting,
- cost/benefit.

Evaluating biometric systems is truly a very important issue that should not be neglected. How to include "true impostors" in the evaluation data and how to make sure that an authentication system that has been tested with a limited population can be scaled and used by millions of users are important problems to consider.

4. Multimodal user authentication

The rational for multimodal user authentication is that no single biometric is generally considered sufficiently accurate, universal, and user-acceptable for any given application. Authentication systems that are robust in natural environments (e.g., in presence of noise and when illumination changes) cannot rely on a single modality. In contrast, multimodal user authentication can provide a more balanced solution to the security and convenience of many applications [29]. However, there is not a clear requirement for the system to be able to adapt to the user needs and conditions, and especially to be able to determine and maintain an acceptable balance. Multimodal user authentication provides a practical and viable approach for overcoming the performance and acceptability barriers to the widespread

Table 1: Pros and cons of multimodal authentication systems.

Pros	Cons
Can overcome the weaknesses of individual biometric identifiers	Can lengthen interaction
Can extend the operation range to a larger target user population	The cost of deployment is generally higher
Can increase the reliability of the decision made by a single biometric system	Integration of multiple modalities is more complex (e.g., score normalization is needed)
Is generally more robust to fraudulent technologies (more difficult to forge)	—
If well designed can improve performance and speed	—

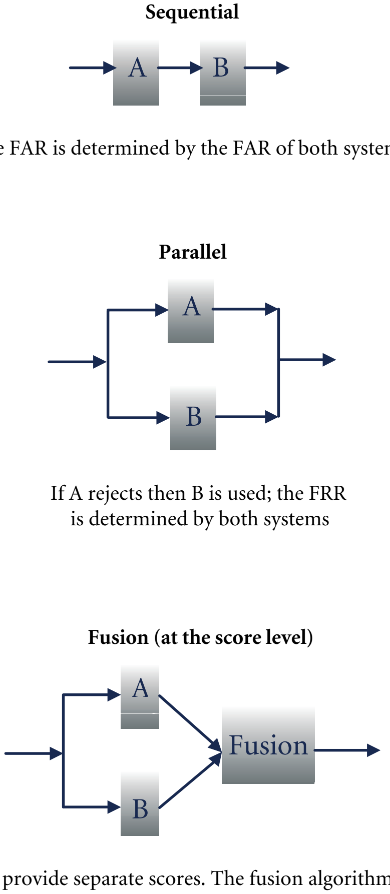
adoption of authentication systems. However, the combination of multimodal biometric modalities is strongly based on a thorough understanding of each of the modalities and the different sensing technologies. A fully successful multimodal fusion can only be obtained through a careful investigation of these technologies and their interaction. Table 1 summarizes the pros and cons.

Multiple modalities correspond to different sources of information. They can be derived from

- the same modality used several times (e.g., combination of two different voiceprint systems or the use of two different fingers for authentication),
- different sources of knowledge (e.g., biometric feature combined with user knowledge),
- any combination of biometric features or any combination of these different methods.

The choice of different modalities is primarily driven by the application requirements. It is important to choose complementary sources of information. For the problem of ASV, it is possible to fuse systems based on low-level features, which are easy to extract but not robust; high-level features are more difficult to extract but more robust [26]. It is also possible to fuse speech features with visual features, such as lip movements, to obtain a system that is more robust in the case of a high acoustic noise [30, 31, 32]. For the problem of AFR, it can be of interest to combine a system based on still intensity images with a system based on infrared (IR) images [33], which is insensitive to illumination variations, or with a system based on range (3D) images [34] which offers an increased robustness with respect to the pose.

The way to combine the modalities is also important [35, 36]. While there is the perception that when a strong test is combined with a weaker test, the resulting decision is averaged, it is important to understand that the performance improvement that can



Sequential

The FAR is determined by the FAR of both systems

Parallel

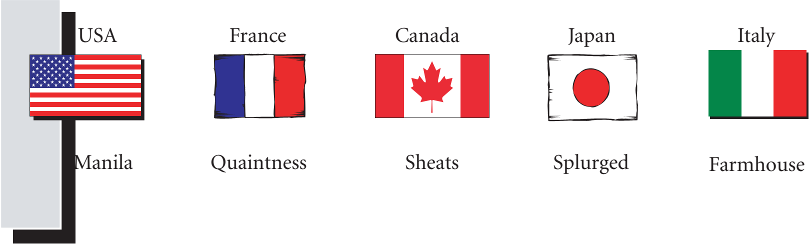
If A rejects then B is used; the FRR is determined by both systems

Fusion (at the score level)

A and B provide separate scores. The fusion algorithm decides. Fusion can produce a very low FAR as well as a very low FRR

Figure 4: Multiple modalities can be combined in different ways.

provide multimodal user authentication comes from a well designed fusion algorithm [37, 38, 39, 40, 41, 42]. As shown in Figure 4, modalities can be combined sequentially, using a parallel architecture, or by means of a fusion algorithm. The fusion algorithm can be applied at the feature level, the decision level, or the score level. The feature level has been so far the best level to understand the correlation between the different modalities. While combining the modalities early on seems like a good idea, it is also more difficult and so far had only limited success. In the example given in Section 5, the modalities (speech and fingerprint) were combined sequentially. In the following paragraphs we would like to emphasize a particular way to combine modalities that is both practical and useful to find a good balance between convenience and security. This is achieved by applying different



In the above example of a challenge response system, random words are extracted from a large word dictionary. Using a rule selected at enrollment time by the user (e.g., select the flag corresponding to the word that contains the letter E and does not contain the letter A, in the above case, ``Splurged”) the user utters the name of the flag (in the above case, ``Japan”) to be authenticated.

Figure 5: Example of challenge response system.

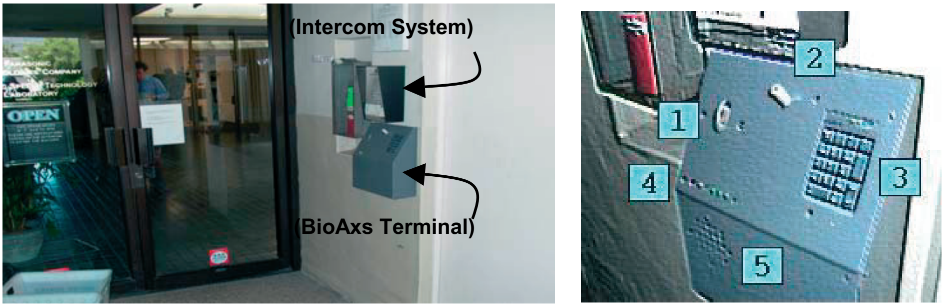
weights to the modalities used. Indeed, it is possible to develop a secure and scalable multimodal user authentication solution by effectively combining biometric features and knowledge information into a system that can be called “challenge-response system.” In such a system, as shown in Figure 5, the knowledge that holds a particular individual is combined with biometric features to provide a highly reliable and scalable system. The advantage of such a solution is that it is easily portable across devices, it is convenient for the users (if the knowledge is not too difficult to remember), and scalable security levels can be achieved.

5. An application example

In this section, a multimodal access terminal for securing the access to a laboratory facility is described [43]. This terminal (BioAxs) hosts two biometric modalities (fingerprint and voiceprint) and one nonbiometric modality (keypad). It was built and installed outside near the main entrance door (first prototype deployed in April 2002) of Panasonic Speech Technology Laboratory. Figure 6 shows a picture of the actual biometric terminal. The terminal is connected to a desktop computer located inside the building via a USB connection.

During the early stage, it became apparent that in the context of this task, a fast and robust interaction model was a necessity. Convenience became therefore a primary concern for success since all employees could always resort to using their key to enter the building.

The access terminal can run in monitoring mode or in user mode. In monitoring mode, the terminal monitors the three sensors in parallel to provide multimodal access control. As explained later in more details the authentication procedure enables single-modality user authentication for fast interaction, and multimodality is used to provide smooth uncertainty recovery. In user mode, the terminal is used for account management or used to run user-dependent commands. In that mode, users must first login by entering their 10-digit account number. Once recognized, authorized users can, for instance, enroll (or re-enroll)



(a)

(b)

Figure 6: (a) The entrance door and the biometric terminal. (b) Picture of the biometric terminal showing (1) the fingerprint scanner, (2) the microphone, (3) the keypad, (4) the LED rack, and (5) the loudspeaker components.

their voiceprint as well as adapt their existing voiceprint. This self-service mode does not require the need for a system administrator. The system is available to all employees and to a selected number of frequent visitors (e.g., employees of United Postal Service).

To provide users with a fast and convenient interaction model, a speaker verification engine (primary modality) was therefore developed based on the following main features.

- *Contact-less activation*. The system monitors the audio channel continuously without the need for explicit activation such as a push-to-talk button, for instance.
- *Far-talking microphone*. Users can either speak while standing by the biometric box or, more conveniently, they can speak to it as they are approaching; the typical operating range is between 1 and 10 feet.
- *Password-dependent voiceprint modeling*. Users can register the voice passphrase of their choice to enter the building. The passphrase is used as an active trigger mechanism that allows people (including registered users) to maintain normal conversations in the vicinity of the box.
- *Password-spotting input mode*. Because the biometric box is located outside the building and is equipped with a far-talking microphone, an input strategy based on automatic endpoint detection was found unreliable in coping with extraneous noises (e.g., street, air-conditioning equipment) and babble noise. A spotting strategy is not affected by endpoint errors. User convenience is therefore increased at the expense of an additional burden on the acceptance/rejection module especially in the case of short passwords (e.g., “California”).

The authentication procedure is primarily unimodal to speed up the door access process but all modalities (keypad, fingerprint, and voiceprint) are available at all times. Upon successful authentication, the entrance door’s contact relay is automatically activated and the name of the verified user is played back along with a series of beeps (from 1 to 5 beeps)

Table 2: Combination of the different modalities for authentication.

	Confidence (1st input)	Confidence (2nd input)	Authentication
Single modality	Low	Low	No
	Low	High	Yes
Multiple modalities	Low	Low	Yes
	Low	High	Yes

Table 3: Usage of the different modalities in real operation.

Modality	Usage	Choice factor	Pros/cons
Voiceprint	95%	Fast and hands-free	Voice masking (loud noise, babble speech)
Fingerprint	5%	Slower	A lot less sensitive to environment
Keypad	<0.01%	Slow and tedious	Always works / must remember access code

indicative of the level of confidence. The multimodal approach helps in the recovery of imperfect matches. This condition occurs when the authentication score is close to the modality's equal error rate. In that case, the security constraints of the helping modality can be reduced without compromising the overall security level, which in the end results in a more robust protocol.

In the case where the user initiates the authentication process by saying his/her voice passphrase, one of three conditions can occur. Based on the authentication score, the system may (1) grant access, (2) deny access, or (3) request additional credentials via another modality. In the latter case, the user can either place his/her finger on the scanner or enter his/her magic key (currently that key corresponds to first digit of user account number) on the keypad. If the credentials are compatible with the hypothesized identity (cross-validation), then access is granted, it is denied otherwise. The user is also allowed to retry by voice but by doing so, the original security settings (i.e., not reduced) must remain in effect.

Single-modality access is in effect during core business hours. During noncore hours such as at night or during weekends multimodal access (i.e., 2 out 3 modalities must be used) is required to enter the building. In that case, the second modality is used to increase the level of security at the expense of a reduction in user convenience. Table 2 indicates how the different modalities were combined and Table 3 summarizes the usage of the three modalities.

This system was installed in April 2002 and was evaluated with an average of 32 users during 19 months of service. In average there were 140 authentications per day. The evaluation of this system revealed that the voice equal error rate was 2.8% with 8% FRR for 0.1% FAR. With the additional credentials (use of additional modalities), 37% of the false rejections could be recovered yielding a combined FRR of 5%. We observed that users tend to say their passwords just well enough to unlock the door. This is natural as there is no extra reward for high voice quality. This explains why the FRR tends to be relatively high.

6. Critical issues linked to application deployment

The most popular markets for biometric systems can be classified in the following categories: forensics, information system/computer network security, physical access, citizen identification, and surveillance. One of the main factors which affect the deployment of biometrics (besides the level of performance) is linked to privacy. Privacy is becoming an increasingly important issue especially for large systems. Consequently issues such as ID management/ID theft, and database management/integrity should not be neglected. It is also important that the biometric system does not store raw data in a database and does not use the biometric data outside the specific purpose. Another issue with privacy is to make sure that it is not possible to recreate the original signal from the stored template. There are ways of ensuring that the biometric template is stored in such a manner that the original biometric cannot be recreated from the template data.

Many additional criteria have to be considered from a privacy point of view: conditions of storage, duration of the data, personal information associated with biometric data, type of population involved in the system, and so forth. Each country has organizations to regulate privacy issues. For example, in France the use of fingerprints for school canteens has been rejected by the CNIL.[1] But a similar system based on hand geometry was accepted. In addition to the criminal connotation of fingerprints, hand geometry features for children vary and then the data has a limited time of validity, so that the risk of future illegal utilization is very low.

Some other applications can use anonymous biometric data going around the privacy issues (i.e., no name or identity information is associated with the stored data). This is the case, for example, of lockers. People may use a fingerprint instead of a key or a code to control a locker. Biometric data are erased once the person comes back and re-opens the locker to retrieve his/her belongings.

Other critical issues to consider when deploying biometric systems include the understanding of consumer expectations and concerns, and the understanding of consumer attitudes towards this technology. Finally, as biometric standards such as BioAPI and CBEFF are becoming more popular, the deployment of biometric systems will become less costly and more efficient.

7. Future directions

There is no doubt that multimodal biometrics and more generally biometric systems will play vital roles in the next generation of authentication systems. However, as was

[1] CNIL stands for “Commission Nationale de l’Informatique et des Libertés.” Founded in 1978, the CNIL is an independent administrative authority protecting privacy and personal data (http://www.cnil.fr).

highlighted in this tutorial there are a number of challenges that need to be tackled. Of course accuracy is still an issue for most deployed systems. Other research areas include

- feature extraction,
- enrollment using a small amount of training data,
- how to combine information (fusion) and make use of the strengths of each modality,
- collection of multimodal and realistic databases (most of the existing databases are unimodal),
- integrating higher level of information (e.g., in the case of speech, prosody, and word/phase usage),
- scalability,
- ease of use,
- privacy concerns,
- establishment of common standards.

Facing these challenges and making progress in these areas will lead to the next generation of biometric systems.

Acknowledgments

The authors would like to acknowledge Philippe Morin from Panasonic Speech Technology Laboratory (now Panasonic Digital Networking Laboratory) for developing the application outlined in Section 5 and the challenge response system shown in Figure 5.

References

[1] J. Wayman, in *National Biometric Test Center Collected Works 1997–2000* (2000), pp. 21–23.
[2] A. Jain, R. Bolle, and S. Pankanti, eds., *Biometrics, Personal Identification in Networked Society* (Kluwer Academic Publishers, 1998).
[3] J. Ashbourn, *Biometrics, Advanced Identity Verification, The Complete Guide* (Springer, 2000).
[4] S. Nanavati, M. Thieme, and R. Nanavati, *Biometrics, Identity Verification in a Networked World* (Wiley Computer Publishing, 2002).
[5] R. Chellappa, C. L. Wilson, and S. Sirohey, Proceedings of the IEEE **83**, 705 (1995).
[6] M. Turk and A. Pentland, J. Cognitive Neuroscience **3**, 71 (1991).
[7] S. Pigeon and L. Vandendorpe, Signal Processing **69**, 59 (1998).
[8] R. Adhami and P. Meenen, IEEE Potentials **20**, 33 (2001).
[9] A. Jain and S. Pankanti, in *Advances in Fingerprint Technology* (Elsevier Science, New York, 2001), 2nd ed.
[10] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzales-Marcos, IEEE PAMI **22**, 1168 (2000).
[11] J. Daugman, in *Biometrics, Personal Identification in Networked Society*, edited by A. Jain, R. Bolle, and S. Pankanti (Kluwer Academic Publishers, 1998), pp. 103–121.
[12] B. Moreno, A. Sanchez, and J. F. Velez, in *IEEE 33rd Annual Int. Carnahan Conf. on Security Technology* (1999), pp. 469–476.
[13] M. Burge and W. Burge, in *Proc. 21st Workshop of the Austrian Association for Pattern Recognition* (Austrian Computer Society, Wien, Hallstatt, Austria, 1997), pp. 275–282.
[14] R. Plamondon and G. Lorette, Pattern Recognition **22**, 107 (1989).

[15] S. Furui, in *Proc. 1st Int. Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA-97)* (1997), pp. 237–251.

[16] D. A. Reynolds and L. P. Heck, in *Tutorial, ICASSP* (Salt Lake City, Utah, 2001).

[17] G. Doddington, in *Eurospeech 2001* (Aalborg, Denmark, 2001), vol. 4, pp. 2521–2524.

[18] O. Thyes, R. Kuhn, P. Nguyen, and J.-C. Junqua, in *ICSLP-2000* (Beijing, China, 2000), vol. 2, pp. 242–245.

[19] *FVC 2004: fingerprint verification competition*, http://bias.csr.unibo.it/fvc2004/.

[20] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone, Evaluation Report (2003).

[21] M. Przybocki and A. Martin, in *The Advent of Biometircs on the Internet, A COST 275 Workshop* (2002).

[22] A. K. Jain, A. Ross, and S. Prabhakar, IEEE Trans. on Circuits and Systems for Video Technology **14** (2004).

[23] J. Wayman, in *National Biometric Test Center Collected Works 1997-2000* (2000), pp. 1–19.

[24] D. Doddington, in *Eurospeech* (2001), vol. 4, pp. 2517–2520.

[25] D. Klusacek, J. Navratil, D. Reynolds, and J. Campbell, in *ICASSP* (2003), vol. 4, pp. 804–807.

[26] D. Reynolds et al., in *ICASSP* (2003), vol. 4, pp. 784–787.

[27] A. J. Mansfield and J. L. Wayman, NPL Report CMSC 14/02 (2002).

[28] P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki, Computer pp. 56–63 (2000).

[29] J.-L. Dugelay, J.-C. Junqua, C. Kotropoulos, R. Kuhn, F. Perronnin, and I. Pitas, in *ICASSP 2002* (2002), pp. IV 4060–IV 4063.

[30] P. Aleksic and A. Katsagellos, in *MMUA* (2003), pp. 80–84.

[31] B. Duc, E. S. Bigun, J. Bigun, G. Maitre, and S. Fischer, Pattern Recognition Letters **18**, 835 (1997).

[32] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, IEEE Trans. Neural Networks **10**, 1065 (1999).

[33] X. Chen, P. Flynn, and K. Bowyer, in *MMUA* (2003), pp. 48–55.

[34] K. Chang, K. Bowyer, and P. Flynn, in *MMUA* (2003), pp. 25–32.

[35] K. Jain, L. Hong, and Y. Kulkarni, Technical Report MSU-CPS-98-32, Department of Computer Science, Michigan State University (1998).

[36] L. Hong and A. Jain, IEEE Trans. Pattern Analysis and Machine Intelligence **20**, 1295 (1998).

[37] R. Brunelli and D. Falavigna, IEEE Trans. Pattern Analysis and Machine Intelligence **17**, 955 (1995).

[38] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, IEEE Trans. Pattern Analysis and Machine Intelligence **20**, 226 (1998).

[39] R. W. Frischholz and U. Dieckmann, IEEE Computer **33**, 64 (2000).

[40] V. Chatzis, A. G. Bors, and I. Pitas, IEEE Trans. Systems, Man and Cybernetics, Part A **29**, 674 (1999).

[41] C. Burges, Knowledge Discovery and Data Mining **2** (1998).

[42] B. Gutschoven and P. Verlinde, in *Fusion '2000* (Paris, 2000).

[43] P. Morin and J.-C. Junqua, in *MMUA* (2003), pp. 19–24.

Florent Perronnin received his Engineering degree in 2000 from the École Nationale Supérieure des Télécommunications, Paris, France. From 2000 to 2001, he was with the Panasonic Speech Technology Laboratory, Santa Barbara, California, first as an intern and then as a Research Engineer, working on speech and speaker recognition. He was then a Ph.D. candidate at the Multimedia Communications Department, Institut Eurécom, Sophia Antipolis, France, focusing his research on automatic face recognition. In 2003, he received a best Student Paper Award at the IEEE International Conference on Image Processing, for the paper "Deformable face mapping for person identification." In 2004, he obtained his Ph.D. degree from the École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. He then joined the Xerox Research Centre Europe, Meylan, France, where he is currently a Research Engineer.

Jean-Claude Junqua received his Engineering degree in 1980 from ENSEM, France in electronics and automation, his Masters and Doctorate degrees in 1981 and 1989, respectively, and the "Habilitation à diriger des recherches" in 1993 from the University of Nancy I, France in the field of computer science. From 1981 to 1986 he was responsible for the computer facilities of CRIN (Research Center in Computer Science of Nancy, France). From 1987 to 1988 he was Visiting Researcher at Panasonic's Speech Technology Laboratory, Santa Barbara, California. In 1989, he joined the Speech Technology Laboratory where he served as a Research Engineer, Group Manager, and Director of the laboratory until March 2004. From April 1992 to August 1993 he was Visiting Researcher at Matsushita Electric Industrial Co., Ltd., Osaka, Japan. Currently he is leading the Panasonic global speech effort at the AV Core Technology Development Center (ACC), Matsushita Electric Industrial Co., Ltd., Osaka, Japan. His interests cover all aspects of automatic speech recognition, speech synthesis, dialogue, multimodal systems, and biometrics. He is the author/coauthor of more than 100 articles and 80 patents in the above areas along with two books entitled *Robustness in Automatic Speech Recognition* and *Robust Speech Recognition in Embedded Systems and PC Applications*. At the beginning of 2001, he also coedited the book *Robustness in Languages and Speech Technology*. He served as a Chairman at several international conferences, coorganized several ISCA/IEEE workshops, and participated in various international scientific committees. Jean-Claude Junqua was a Tutorial Speaker for several ESCA/IEEE workshops, ICASSP'1999, ICSLP'2002, and ICASSP'2004. He was an Associate Editor of the IEEE Transactions on Speech and Audio Processing and he is currently on the Editorial Board of the Speech Communication Journal and the ACM magazine, Computers in Entertainment, while serving on the Speech Technical Committeee of IEEE.

Jean-Luc Dugelay received the Ph.D. degree in computer science in 1992 from the University of Rennes. Doctoral research was carried out from 1989 to 1992 at the France Telecom Research Laboratory in Rennes (formerly CNET - CCETT). He then joined the Institut Eurécom, Sophia Antipolis, where he is currently a Professor in the Department of Multimedia Communications. His research interests are in the area of multimedia signal processing and communications; including security imaging (i.e., watermarking and biometrics), facial image analysis, and talking heads. He is an author or coauthor of more than 65 publications that have appeared as journal papers or proceeding articles, 3 book chapters, and 3 international patents. He gave several tutorials on digital watermarking (coauthored with F. Petitcolas from Microsoft Research), and biometrics (coauthored with J.-C. Junqua from Panasonic Research) at major conferences. He has been an invited speaker and/or member of the program committee of several scientific conferences and workshops. He was Technical Cochair and organizer of the fourth workshop on Multimedia Signal Processing (Cannes, October 2001), and coorganizer of the workshop on Multimodal User Authentication (Santa Barbara, December 2003). His group is involved in several national and European projects related to biometrics. Jean-Luc Dugelay is a Senior Member of the IEEE Signal Processing Society, and is currently an Associate Editor for the EURASIP Journal on Applied Signal Processing.