Data Fusion at Different Levels

Marcos Faundez-Zanuy

Escola Universitària Politècnica de Mataró
Avda. Puig i Cadafalch 101-111, 08303 MATARÓ (BARCELONA)
faundez@eupmt.es

Abstract. This paper summarizes the main characteristics of data fusion at different levels (sensor, features, scores and decisions). Although it is presented in the framework of biometric applications it is general for all the pattern recognition applications because this presentation is focused in the main blocks of a general pattern recognition system. Thus, the application in mind will imply a different sensor, feature extractor, classifier and decision maker but data fusion will be performed in a similar way.

1 Introduction

There has been a paradigm shift in the approach to solving pattern recognition problems [1-2]: Instead of looking for the best set of features and the best classifier, now we look for the best set of classifiers and then the best combination method. To solve really hard problems, we will have to use several different representations. It is time to stop arguing over which type of pattern classification technique is best because that depends on our context and goal. Instead we should work at a higher level of organization and discover how to build managerial systems to exploit the different virtues and evade the different limitations of each of these ways of comparing things.

There are several scientific fields where data fusion is performed. Some examples [3] are the following:

- Weather forecasting: forecasting systems rely on the evidence provided by diverse sources of information such as geostationary meteorological satellites, weather balloons, ground stations, radars, etc.
- Robot navigation: a robot is typically fitted with a variety of sound, light, image, proximity etc. sensors
- Land mine detection: several types of sensor technologies are being used to detect buried land mines: electromagnetic induction, ground penetrating radar, infra-red imaging, chemical detectors, etc.

The general scheme of a pattern recognition system consists of four main blocks, as shown in figure 1. Each block corresponds to one possible data fusion level. Data fusion combines sources of multiple evidences. In the case of biometrics [4], the following sources can be identified:

- Multi-sample (for instance, several snapshots of a face)
- Multi-modal (example: face and speech)

A. Esposito et al. (Eds.): Multimodal Signals, LNAI 5398, pp. 94–103, 2009.

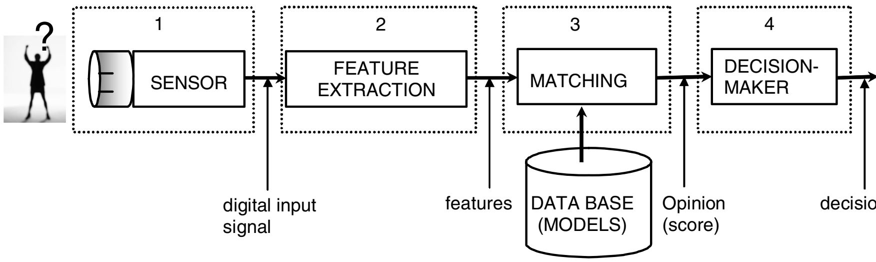

Fig. 1. General scheme of a biometric system

- Multi-sensor (for instance, two microphones for rec. speech)
- Multi-algorithm (for instance, HMM and SVM)
- Multi-instance (for instance, left and right irises)

In all the cases, the system can be classified as:

a) Unimodal biometric system: it relies on a single biometric characteristic.
b) Multimodal biometric system: it uses multiple biometric characteristics, like voice plus fingerprint or face plus iris.

Usually the unimodal systems are easier to install, the computational burden is typically smaller, they are easier to use, and cheaper, because just one sensor (or several sensors of the same kind) are needed. On the other hand, a multimodal system can overcome the limitations of a single biometric characteristic.

2 Data Fusion Levels

Considering the main blocks plotted in figure 1, the following levels can be defined:

Sensor Level

In this level, the digital input signal is the result of sensing the same biometric characteristic with two or more sensors. Thus, it is related to unimodal biometrics. Figure 2 shows an example of sensor fusion that consists of sensing a speech signal simultaneously with two different microphones for a speaker recognition application [5]. The combination of the input signals can provide noise cancellation, blind source separation, etc.

Another example is face recognition using multiple cameras that are used to acquire frontal and profile images in order to obtain a three dimensional face model, which is used for feature extraction.

Although this fusion level is useful in several scenarios, it is not the most usual one.

Feature Level

This level can apply to the extraction of different features over a single biometric signal (unimodal system) and the combination of feature levels extracted from different

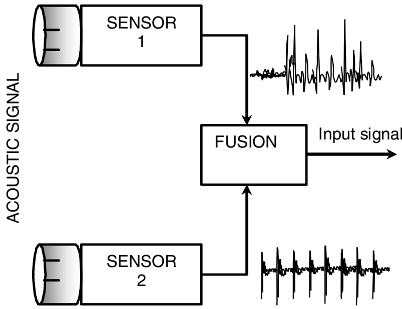

Fig. 2. Example of sensor fusion for speech signals. This block should replace block number 1 in fig. 1.

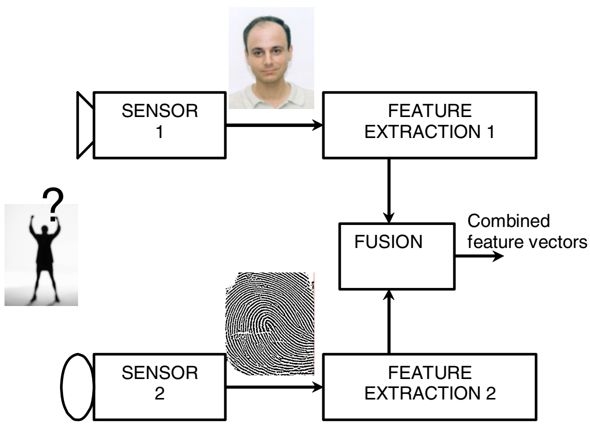

Fig. 3. Example of feature fusion. This block should replace blocks number 1 and 2 in figure 1.

biometric characteristics (multimodal system). An example of a unimodal system is the combination of instantaneous and transitional information for speaker recognition.

Figure 3 shows an example that consists of a combination of face and fingerprint at the feature level.

This combination strategy is usually done by a concatenation of the feature vectors extracted by each feature extractor. This yields an extended size vector set.

Some drawbacks of this fusion approach are:

- There is little control over the contribution of each vector component on the final result, and the augmented feature space can imply a more difficult classifier design, the need for more training and testing data, etc.
- Both feature extractors should provide identical vector rates. This could not be a problem for the combination of face and fingerprint, because one vector per acquisition can be obtained. However, it can be a problem for combining voice with another biometric characteristic, due to the high number of vectors that depend on the test sentence length (if both feature extractors do not provide the same amount of vectors per trial, it is not possible to concatenate the vectors extracted by each feature extractor)

Although it is a common belief that the earlier the combination is done, the better is the result achieved, state-of-the-art data fusion relies mainly on the opinion and decision levels.

Opinion Level

This kind of fusion is also known as confidence level. It consists of the combination of the scores provided by each matcher. The matcher just provides a distance measure or a similarity measure between the input features and the models stored on the database.

It is possible to combine several classifiers working with the same biometric characteristic (unimodal systems) or to combine different ones. Figure 4 shows an example of multimodal combination of face and iris.

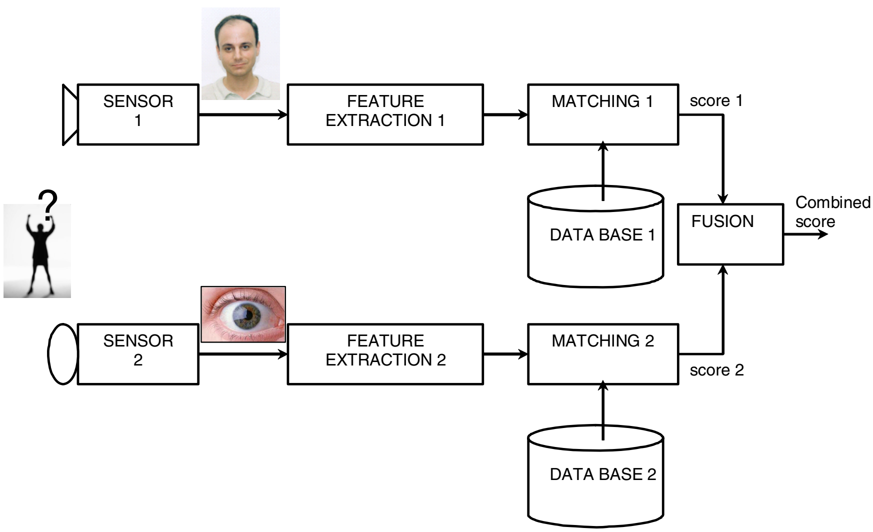

Fig. 4. Example of opinion fusion. This block should replace blocks number 1, 2, and 3 in figure 1.

Before opinion fusion, normalization must be done. For instance, if the measures of the first classifier are similarity measures that lie on the [0, 1] range, and the measures of the second classifier are distance measures that range on [0, 100] two normalizations must be done:

1. The similarity measures must be converted into distance measures (or vice versa).
2. The location and scale parameters of the similarity scores from the individual classifiers must be shifted to a common range. Although several approaches to score normalization exist, this is still an unsolved problem. A given normalization strategy will not be the optimal for all the scenarios. One possible normalization consists of using a sigmoid function: $o_i' = \frac{1}{1+e^{-k_i}}$, with $k_i = \frac{o_i - (m_i - 2\sigma_i)}{2\sigma_i}$

where:

$o_i' \in [0,1]$, o_i is the initial opinion of the classifier i.

m_i, σ_i are the mean and standard deviation of the classifier i opinions obtained with data belonging to genuine users (scores obtained comparing feature vectors belonging to the same user as the model).

After the normalization procedure, several combination schemes can be applied [6].

Figure 5 shows an example of two speaker recognition system, the first one is based on Covariance matrices and the second one in Vector Quantization.

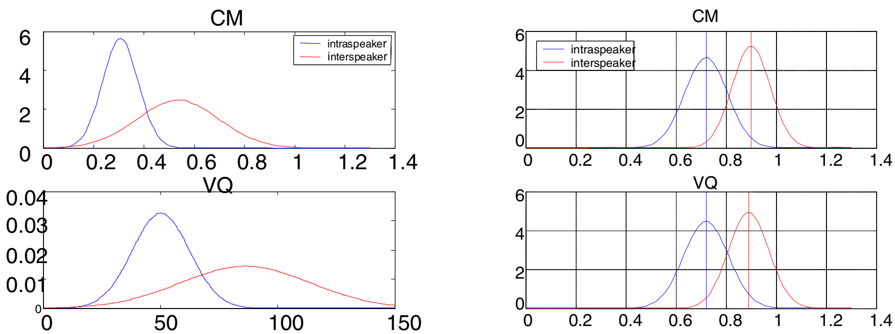

Fig. 5. Example of histograms for intra and interdistances for two speaker recognition systems Covariance matrices (CM) and Vector Quantization (VQ), before (on the left) and after normalization (on the right)

The combination strategies can be classified into three main groups:

1. Fixed rules: All the classifiers have the same relevance. An example is the sum of the outputs of the classifiers. That is: let o_1 and o_2 be the outputs of classifiers number 1 and 2 respectively. For example, a fixed combination rule yields the combined output $O = (o_1 + o_2)/2$
2. Trained rules: Some classifiers should have more relevance on the final result. This is achieved by means of some weighting factors that are computed using a training sequence. That is: $O = \omega_1 o_1 + \omega_2 o_2 = \omega_1 o_1 + (1 - \omega_1) o_2$. Figure 6 shows an example of a trained rule that consists of the combination of two different classifiers for speech recognition. It is interesting to observe that for $\omega_1 = 1$ (83.8% identification rate) just the first classifier is considered, while for $\omega_1 = 0$ (79.2% identification rate) just the second classifier has relevance. For intermediate values, higher identification rates are achieved (84.8%).
3. Adaptive rules: The relevance of each classifier depends on the instant time. This is interesting for variable environments. That is: $O = \omega_1(t) o_1 + (1 - \omega_1(t)) o_2$. For instance, a system that combines speech and face can detect those situations where the background noise increases and then reduce the speech classifier weight. Similarly, the face classifier weight is decreased when the illumination degrades or there is no evidence that a frontal face is present.

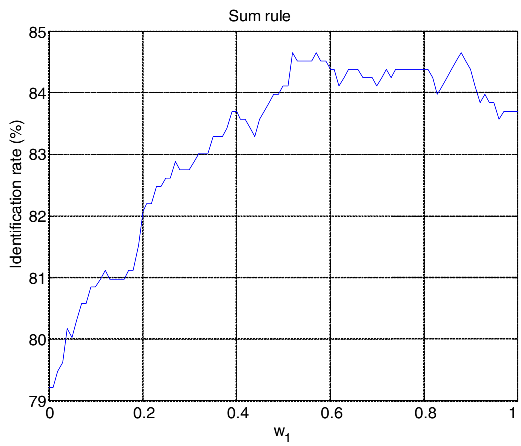

Fig. 6. Example of trained rule for opinion fusion. It combines two different speaker recognition classifiers.

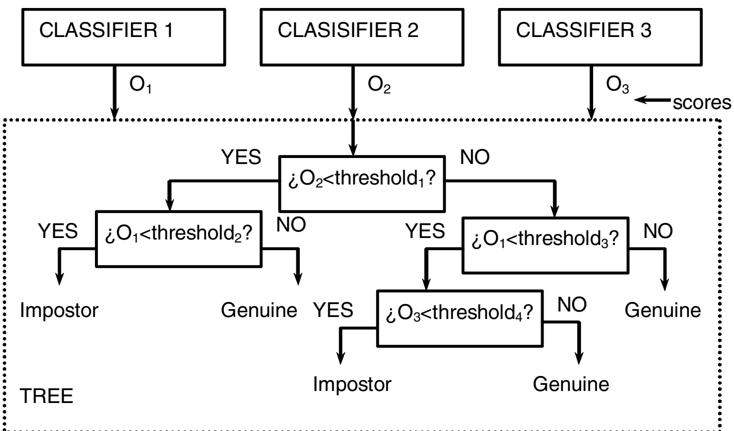

Fig. 7. Example of opinion fusion using classifier trees

The most popular combination schemes are:

1. Weighted sum: $O_j = \sum_{i=1}^{N} \omega_i o_{ij}$. A particular case would be the arithmetic mean. for instance, for two classifiers combination assigning equal weight to both of them, $O = (o_1 + o_2)/2$.
2. Weighted product: $O_j = \prod_{i=1}^{N} (o_{ij})^{\omega_i}$. A particular case would be the geometric mean. For instance, for a two classifiers combination assigning equal weight to both of them, $O = \sqrt{o_1 \times o_2} = (o_1)^{\frac{1}{2}} \times (o_2)^{\frac{1}{2}}$

3. Decision trees: it is based on if-then-else sentences. Figure 7 shows an example of data fusion using a decision tree.

Figure 8 shows the confusion matrices of the CM and VQ classifiers of figures 5 and 6. Each element (k, r) of the confusion matrix represents the number of instances in the test data set where a pattern whose true class label k is assigned to a class r. If there were no errors all the nonzero elements would be in the diagonal ($k = r$). It is interesting to observe that the classification errors performed by both classifiers are not exactly the same. That is, they make different errors. This is the key point for obtaining better results after combination.

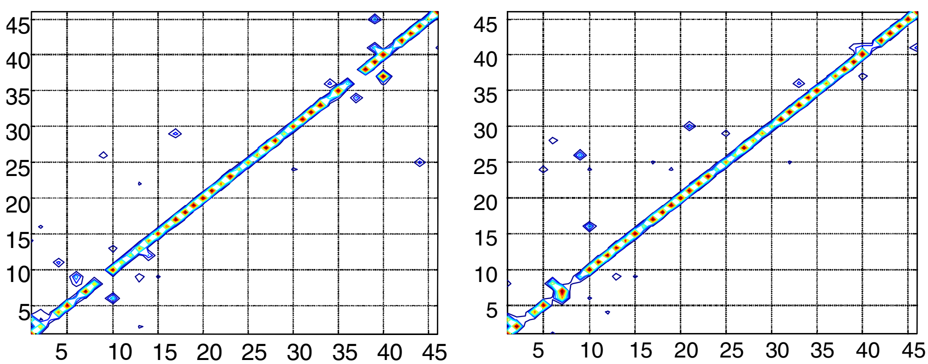

Fig. 8. Confusion matrices for CM and VQ classifiers of figure 5 and 6

Decision Level

At this level, each classifier provides a decision. On verification applications it is an accepted / rejected decision. On identification systems it is the identified person or a ranked list with the most probable person on its top. In this last case, the Borda count method [7] can be used for combining the classifiers' outputs. This approach overcomes the scores normalization that was mandatory for the opinion fusion level. Figure 9 shows an example of the Borda count. The Borda count assigns a score that is equal to the number of classes that are ranked below the given class.

One problem that appears with decision level fusion is the possibility of ties. For verification applications, at least three classifiers are needed (at least two of them will agree and there is no tie), but for identification scenarios the number of classifiers should be higher than the number of classes. This is not a realistic situation, so this combination level is usually applied to verification scenarios.

An important combination scheme at the decision level is the serial and parallel combination, also known as "AND" and "OR" combinations. Figure 10 shows the block diagram. In the first case, a positive verification must be achieved in both systems, while access is achieved in the second one if the user is accepted by one of the systems. If each system is characterized by its False Acceptance Rate (FAR) and False Rejection Rate (FRR) FAR_1 , FRR_1, FAR_2 , FRR_2 ,the combined systems provide:

$FAR_{AND} = FAR_1 \times FAR_2$
$FRR_{AND} = FRR_1 + (1–FRR_1) \times FRR_2$
$FAR_{OR} = FAR_1 + (1–FAR_1) \times FAR_2$
$FRR_{OR} = FRR_1 \times FRR_2$

While serial combination (AND) improves security (FAR is reduced), parallel combination (OR) improves user convenience (FRR is reduced).

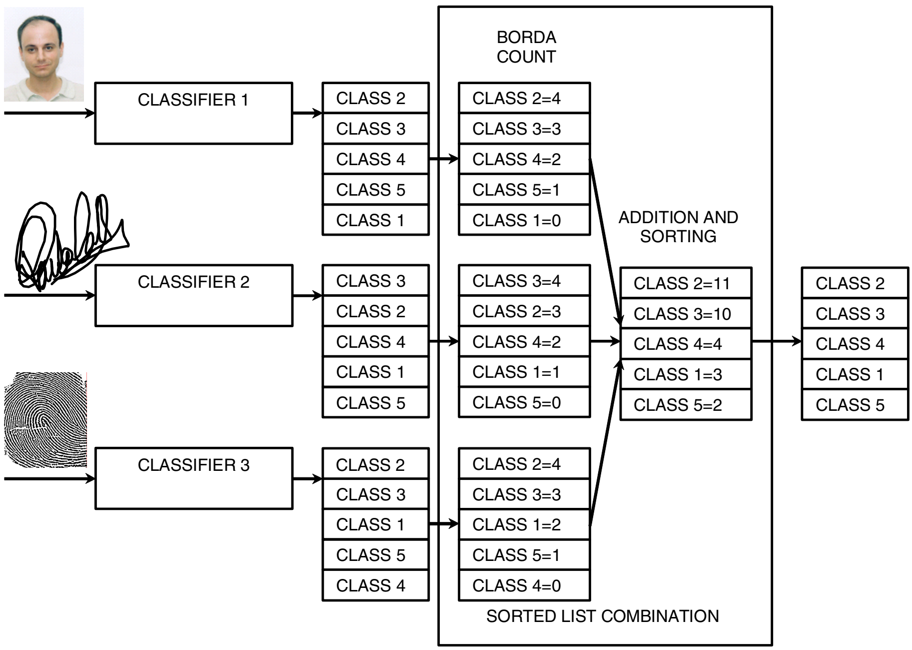

Fig. 9. Example of decision level fusion. It combines face, signature and fingerprint by means of the Borda counts.

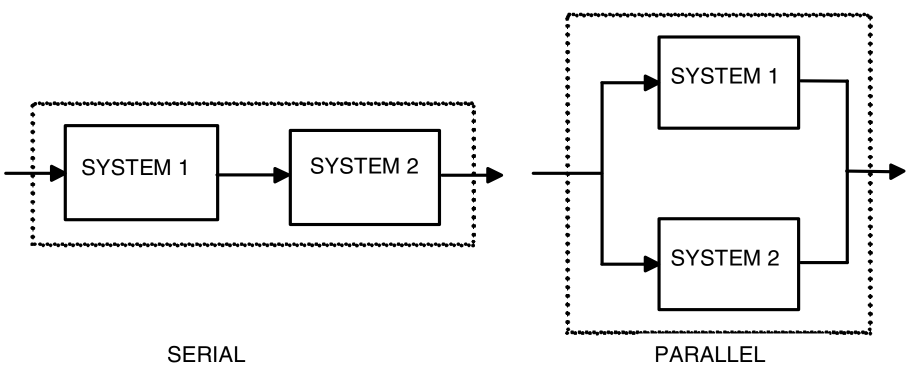

Fig. 10. Serial and parallel decision level combinations

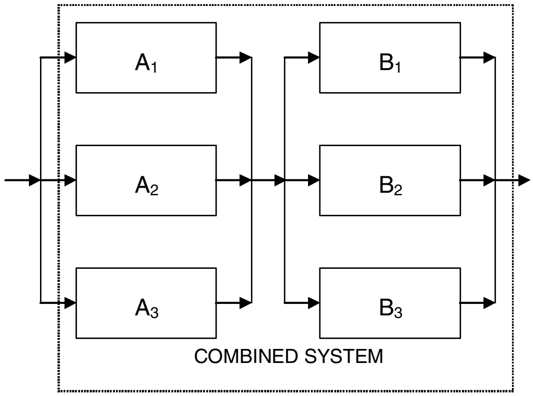

Fig. 11. Example of simultaneous combination in serial and parallel for improving both rates (FAR and FRR). It consists of the serial combination of two systems (A and B) offering three trials on each of them.

Simultaneously combining serial and parallel systems, it is possible to improve both rates. For instance, [8] reports the combination of two different biometric systems offering three trials in each one (similar to the PIN keystroke on ATM cashiers). In this case, if each system on its own yields a 1% False Acceptance Ratio (FAR) and 1% False Rejection Ratio (FRR), the combined system yields FAR=0.0882% and FRR=0.0002. (see figure 11). In this case, for the three parallel blocks we achieve:

$$\begin{cases} TFA(\%) = \left[TFA_1 + (1-TFA_1)\times TFA_2 + \left(1-\left(TFA_1 + (1-TFA_1)\times TFA_2\right)\right)\times TFA_3\right]\times 100 = 2.9\% \\ TFR(\%) = [TFR_1 \times TFR_2 \times TFR_3]\times 100 = 0.0001\% \end{cases}$$

And then, for the resulting two serial blocs, we achieve:

$$\begin{cases} TFA_{Eq}(\%) = [TFA_A \times TFA_B]\times 100 = 0.0882\% \\ TFR_{Eq}(\%) = \left[TFR_A + (1-TFR_A)\times TFR_B\right]\times 100 = 0.0002\% \end{cases}$$

Thus, the improvement is evident because the mean error is:

$$\frac{1}{2}\left(TFA_{Eq} + TFR_{Eq}\right) = 0.0442\%$$

while initially it was $\frac{1}{2}\left(TFA_{Eq} + TFR_{Eq}\right) = 1\%$.

3 Conclusions

In this paper we have summarized the four main blocks for data fusion in a pattern recognition system. Although this presentation is related to biometric recognition it can be easily generalized to other pattern recognition applications.

Data fusion is not a solved problem, and for instance, our recent work [9] deals with the combination of a large set of sources of information. In this scenario a brute force method for weighting computation is not possible and a Maximum likelihood Linear Programming Data Fusion for Speaker Recognition is proposed.

Acknowledgement

This work has been supported by FEDER and the Spanish grant MCYT TIC2003-08382-C05-02.

References

1. Ho, T.K.: Multiple classifier combination: lessons and next steps. In: Bunke, H., Kandel, A. (eds.) Hybrid methods in pattern recognition. Machine perception and artificial intelligence, vol. 47, pp. 171–198. World Scientific, Singapore (2002)
2. Minsky, M.: logical versus analogical or symbolic versus connectionist or neat versus scruffy. AI magazine 12, 34–51 (1991)
3. Ross, A.A., Nandakumar, K., Jain, A.K.: Handbook of multibiometrics. Springer, Heidelberg (2006)
4. Faundez-Zanuy, M.: Biometric security technology. IEEE Aerospace and Electronic Systems Magazine 21(6), 15–26 (2006)
5. Faundez-Zanuy, M., Monte-Moreno, E.: State-of-the-art in speaker recognition. IEEE Aerospace and Electronic Systems Magazine 20(5), 7–12 (2005)
6. Faundez-Zanuy, M.: Data fusion in biometrics. IEEE Aerospace and Electronic Systems Magazine 20(1), 34–38 (2005)
7. Ho, T.K., Hull, J.J., Srihari, S.N.: Decision combination in multiple classifier systems. IEEE Trans. On Pattern analysis and machine intelligence 16(1), 66–75 (1994)
8. Willems, M., Forret, P.: Layered Biometric Verification. White paper, Keyware Technologies (December 1997)
9. Monte-Moreno, E., Chetouani, M., Faundez-Zanuy, M., Sole-Casals, J.: Maximum likelihood Linear Programming Data Fusion for Speaker Recognition. Speech Communication. Elsevier, Amsterdam (in press), doi:10.1016/j.specom.2008.05.009