SPEECH AND SIGNATURE BASED MULTI-MODAL BIOMETRICS VERIFICATION

MOHAMED SOLTANE
Electronics Department, Faculty of engineering science
Badji Mokhtar University of Annaba, 23000 Annaba - Algeria
xor99@hotmail.com

Abstract: This paper describes a combined behavioral techniques based on speech and signature biometrics modalities. Fusion of multiple biometric modalities for human verification performance improvement has received considerable attention. Multi-biometric systems, which consolidate information from multiple biometric sources, are gaining popularity because they are able to overcome limitations such as non-universality, noisy sensor data, large intra-user variations and susceptibility to spoof attacks that are commonly encountered in mono modal biometric systems. Soft decision level fusion based Gaussian mixture models (GMM), in which the (EM), (GEM) and (FJ) algorithms for estimating the parameters of the mixture model and the number of mixture components have been compared. The test performance of the fusion, EER=0.0 % for "EM" and "FJ", EER=0.02 % for "GEM", show that the combined behavioral information scheme is more robust and have a discriminating power, which can be explored for identity authentication.

Key words: *Biometric authentication, behavioral biometrics, speech analysis, signature verification, soft decision fusion, Gaussian Mixture Modal, EM, GEM and FJ.*

I. INTRODUCTION

BIOMETRIC is a Greek composite word stemming from the synthesis of bio and metric, meaning life measurement. In this context, the science of biometrics is concerned with the accurate measurement of unique biological characteristics of an individual in order to securely identify them to a computer or other electronic system. Biological characteristics measured usually include fingerprints, voice patterns, retinal and iris scans, face patterns, and even the chemical composition of an individual's DNA [1]. Biometrics authentication (BA) (*Am I whom I claim I am?*) involves confirming or denying a person's *claimed identity* based on his/her physiological or behavioral characteristics [2]. BA is becoming an important alternative to traditional authentication methods such as keys (“something one has", i.e., by possession) or PIN numbers (“something one knows", i.e., by knowledge) because it is essentially “who one is", i.e., by biometric information. Therefore, it is not susceptible to misplacement or forgetfulness [3]. These biometric systems for personal authentication and identification are based upon physiological or behavioral features which are typically distinctive, although time varying, such as fingerprints, hand geometry, face, voice, lip movement, gait, and iris patterns. Multi-biometric systems, which consolidate information from multiple biometric sources, are gaining popularity because they are able to overcome limitations such as non-universality, noisy sensor data, large intra-user variations and susceptibility to spoof attacks that are commonly encountered in mono-biometric systems.

Some works based on multi-modal biometric identity verification systems has been reported in literature. **M. Fuentes et al. [4]** describe two biometrics identity verification systems relying on Hidden Markov Models (HMMs): one for online signature verification and the other one for speaker verification. These two systems are first tested separately, then the scores of each HMM expert have been fused together by different methods. A Support Vector Machine scheme has been shown to improve significantly the results. **A. Perez-Hernandez et al. [5]** propose a simple adaptive off-line signature recognition method based on the feature analysis of extracted significant strokes for a given signature. Their system correctly decides on the majority of tested patterns, which include both simple and skilled forgeries. Experimental results have showed a good trade-off between response time and reasonable recognition accuracy. **Hugo Gamboa et al. [6]** describe a new behavioral biometric technique based on human computer interaction. They developed a system that captures the user interaction via a pointing device, and uses this behavioral information to verify the identity of an individual. Using statistical pattern recognition techniques, they developed a sequential classifier that processes user interaction, according to which the user identity is considered genuine if a predefined accuracy level is achieved, and the user is classified as an impostor otherwise. Two statistical models for the features were tested, namely Parzen density estimation and a uni-modal distribution. The system was tested with different numbers of users in order to evaluate the scalability of the proposal. Experimental results showed that the normal user interaction with the computer via a pointing device entails behavioral information with discriminating power that can be explored for identity authentication. **Ibrahim S. I. ABUHAIBA [7]** presents a simple and effective signature verification method that depends only on the raw binary pixel intensities and avoids using complex sets of features. The method looks at the signature verification problem as a graph matching problem. The method is tested using genuine and forgery signatures produced by five subjects. An equal error rate of 26.7% and 5.6% was achieved for skilled and random forgeries, respectively. A

Manuscript received September 22, 2010; revised November 05, 2010

positive property of the algorithm is that the false acceptance rate of random forgeries vanishes at the point of equal false rejection and skilled forgery false acceptance rates. **Ben-Yacoub et al. [8]** evaluated five binary classifiers on combinations of face and voice modalities (XM2VTS database). They found that (i) a support vector machine and bayesian classifier achieved almost the same performances; and (ii) both outperformed Fisher's linear discriminent, a C4.5 decision tree, and a multilayer perceptron. **Korves et al. [9]** compared various parametric techniques on the BSSR1 dataset. That study showed that the Best Linear technique performed consistently well, in sharp contrast to many alternative parametric techniques, including simple sum of z-scores, Fisher's linear discriminant analysis, and an implementation of sum of probabilities based on a normal (Gaussian) assumption.

Multi-biometric systems provide a variety of advantages against traditional biometric systems and are able to encounter the performance requirements of various applications [10]. The problem of non-universality is addressed, since sufficient population coverage can be ensured by a multiple traits. Furthermore, multi-biometric systems can facilitate the indexing of large-scale databases, can address the problem of noisy data and provide anti-spoofing measures by making it difficult for an impostor to spoof multiple biometric traits of a legitimate enroll individual.

In this paper a multi-modal biometric verification system based on combined behavioral speech-signature modalities is described. In multimodal systems, complementary input modalities provide the system with non-redundant information whereas redundant input modalities allow increasing both the accuracy of the fused information by reducing overall uncertainty and the reliability of the system in case of noisy information from a single modality. Information in one modality may be used to disambiguate information in the other ones. The enhancement of precision and reliability is the potential result of integrating modalities and/or measurements sensed by multiple sensors [11].

II. VERIFICATION TRAITS

A. Speech Analysis and Feature Extraction

Gaussian Mixture Models (GMMs), is the main tool used in text-independent speaker verification, in which can be trained using the Expectation Maximization (EM) algorithm [12]. In this work the speech modality, is authenticated with a multi-lingual text-independent speaker verification system. The speech trait is comprised of two main components as shown in figure 1: speech feature extraction and a Gaussian Mixture Model (GMM) classifier. The speech signal is analyzed on a frame by frame basis, with a typical frame length of 20 ms and a frame advance of 10 ms [13]. For each frame, a dimensional feature vector is extracted, the discrete Fourier spectrum is obtained via a fast Fourier transform from which magnitude squared spectrum is computed and put it through a bank of filters. The critical band warping is done following an approximation to the Mel-frequency scale which is linear up to 1000 Hz and logarithmic above 1000 Hz. The Mel-scale cepstral coefficients are computed from the outputs of the filter bank [14]. The state of the art speech feature extraction schemes (Mel frequency cepstral coefficients (MFCC) is based on auditory processing on the spectrum of speech signal and cepstral representation of the resulting features [15]. One of the powerful properties of cepstrum is the fact that any periodicities, or repeated patterns, in a spectrum will be mapped to one or two specific components in the cepstrum. If a spectrum contains several harmonic series, they will be separated in a way similar to the way the spectrum separates repetitive time patterns in the waveform. The description of the different steps to exhibit features characteristics of an audio sample with MFCC is showed in figure 2.

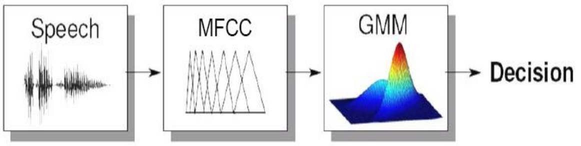


Figure 1. Acoustic Speech Analysis.

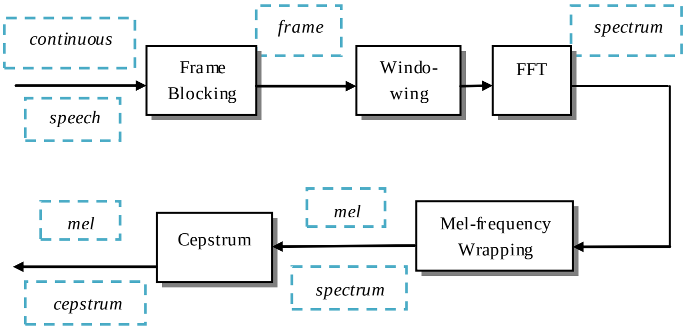


Figure 2. MFCC calculation Block diagram [14].

The distribution of feature vectors for each person is modeled by a GMM. The parameters of the Gaussian mixture probability density function are estimated using three different estimation algorithms. The Expectation Maximization (EM) algorithm [16], Greedy algorithm (GEM) [16] and Figueiredo-Jain (FJ) algorithm [16].

Given a claim for person ***C's*** identity and a set of feature vectors $X = \{\vec{x}_i\}_{i=1}^{Nv}$ supporting the claim, the average log likelihood of the claimant being the true claimant is calculated using:

$$\mathcal{L}(X|\lambda_C) = \frac{1}{N_V}\sum_{i=1}^{N_V} \log p(\vec{x}_i|\lambda_C) \tag{1}$$

where $p(\vec{x}|\lambda) = \sum_{j=1}^{N_M} m_j\, \mathcal{N}\left(\vec{x}\,;\overrightarrow{\mu_j};\sum_j\right)$ (2)

and $\lambda = \left\{m_j\ ,\ \overrightarrow{\mu_j}\ ,\sum_j\right\}_{j=1}^{N_M}$ (3)

Here λ_C is the model for person C. N_M is the number of mixtures, m_j is the weight for mixture j (with constraint $\sum_{j=1}^{N_M} m_j = 1$), and $\mathcal{N}(\vec{x}\,;\ \vec{\mu},\sum)$ is a multi-variate Gaussian function with mean $\vec{\mu}$ and diagonal covariance matrix $\sum$. Given a set $\{\lambda_b\}_{b=1}^{B}$ of ***B*** background person models for person ***C***, the average log likelihood of the claimant being an impostor is found using:

$$\mathcal{L}\left(X\middle|\lambda_{\bar{C}}\right) = \log\left[\frac{1}{B}\sum_{b=1}^{B} \exp \mathcal{L}(X|\lambda_b)\right] \tag{4}$$

The set of background person models is found using the method described in [17]. An opinion on the claim is found using:

$$o = \mathcal{L}(X|\lambda_C) - \mathcal{L}(X|\lambda_{\bar{C}}) \qquad (5)$$

The opinion reflects the likelihood that a given claimant is the true claimant (i.e., a low opinion suggests that the claimant is an impostor, while a high opinion suggests that the claimant is the true claimant).

B. Signature verification systems

Handwritten signature is one of the first accepted civilian and forensic biometric identification technique in our society [7]. Human verification is normally very accurate in identifying genuine signatures. A signature verification system must be able to detect forgeries and at the same time reduce rejection of genuine signatures. The signature verification problem can be classified into categories: offline and online. Offline signature verification does not use dynamic information that is used extensively in online signature verification systems. This paper investigates the problem of on-line signature verification. The problem of on-line signature verification has been faced by taking into account three different types of forgeries: random forgeries, produced without knowing either the name of the signer or the shape of his signature; simple forgeries, produced knowing the name of the signer but without having an example of his signature; and skilled forgeries, produced by people who, looking at an original instance of the signature, attempt to imitate it as closely as possible.

Figure 3. Wacom Graphire3 digitizing TabletPC.

1) Feature Extraction: The coordinate trajectories (x_n, y_n) and pressure signal p_n are the components of the unprocessed feature vectors $u_n = [x_n, y_n, p_n]^T$ extracted from the signature signal [18], where n =1,...,N_s and N_s is the duration of the signature in time samples. Signature trajectories are then pre-processed by subtracting the centre of mass followed by rotation alignment based on the average path tangent angle. An extended set of discrete-time functions are derived from the pre-processed trajectories consisting of sample estimations of various dynamic properties. As s result, the parameterised signature O consists in the sequence of feature vectors $o_n = [x_n, y_n, p_n, \theta_n, v_n, \dot{x}_n, \dot{y}_n]^T$, n =1,...,$N_s$, where the upper dot notation represents an approximation to the first order time derivative and θ *and* v stand respectively for path tangent angle, path velocity magnitude.

$$v_i = \sqrt{\dot{x}_i^2 + \dot{y}_i^2} \quad \text{and} \quad \theta_i = arctan(\dot{y}_i, \dot{x}_i)$$
$$\text{and} \quad \dot{x}_i = x_i - x_{i-1} \quad and \quad \dot{y}_i = y_i - y_{i-1}$$

A whitening linear transformation is finally applied to each discrete-time function so as to obtain zero mean and unit standard deviation function values. Seven dimensional feature vectors are used for GMM processing described in the following section. Figure 5 shows x-, y-, p- and velocity signals of an example signature.

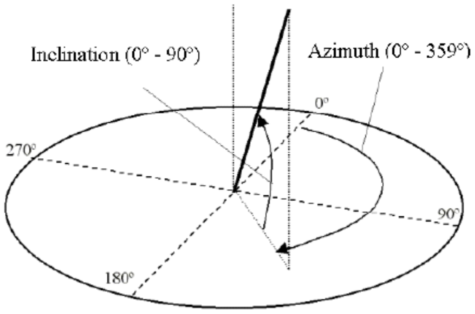


Figure 4. Azimuth and inclination angles of the pen respect to the plane of the graphic card GD-0405U from Wacom Graphire3 digitizing TabletPC.

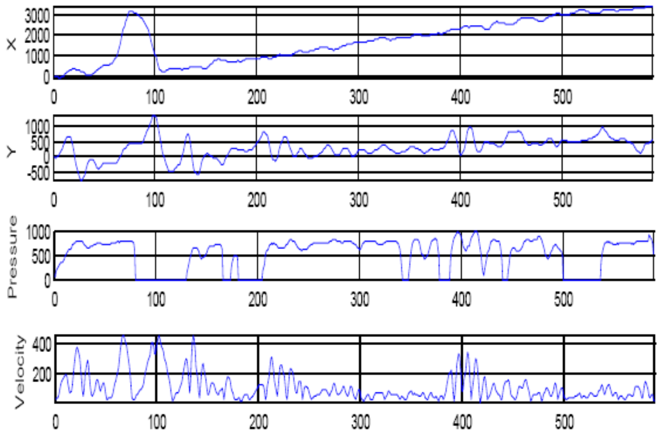


Figure 5. Signals (x-, y- position, pen pressure and velocity) of one signature fragment.

III. MULTIMODAL BIOMETRIC DECISION FUSION METHODS

The process of biometric user authentication can be outlined by the following steps [19]: a) acquisition of raw data, b) extraction of features from these raw data, c) computing a score for the similarity or dissimilarity between these features and a previously given set of reference features and d) classification with respect to the score, using a threshold. The results of the decision processing steps are *true* or *false* (or *accept/reject*) for verification purposes or the user identity for identification scenarios.

The fusion of different signals can be performed 1) at the raw data or the feature level, 2) at the score level or 3) at the decision level. These different approaches have advantages and disadvantages. For *raw data* or *feature level* fusion, the basis data have to be compatible for all modalities and a common matching algorithm (processing step c) must be used. If these conditions are met, the separate feature vectors of the modalities easily could be concatenated into a single new vector. This level of fusion has the advantage that only one algorithm for further processing steps is necessary instead of one for each modality. Another advantage of fusing at this early stage of processing is that no information is lost by previous processing steps. The main disadvantage is the demand of compatibility of the different raw data of features. The fusion at *score level* is performed by computing a similarity or dissimilarity (distance) score for each single modality. For joining of these different scores, normalization should be done. The straightforward and most rigid approach for fusion is the decision level. Here, each biometric modality results in its own decision; in case of a verification scenario this is a set of *true* and *false*. From this set a kind of voting (majority decision) or a logical *"AND"* or logical *"OR"* decision can be computed. This level of fusion is the least powerful, due to the absence of much information. On the other hand, the advantage of this fusion strategy is the easiness and the guaranteed availability of all single modality decision results. In practice, score level fusion is the best-researched approach, which appears to result in better improvements of recognition accuracy as compared to the other strategies.

A. Theoretical Analysis for Decision Level Fusion

The fusion scheme using these two modalities is denoted by S. Verification system based only on speech is denoted by S_1, while on signature by S_2 [20]. If Γ is an algorithm, then the task is to find which acts on independent sources so that the output is maximized. It can be written as:

$$\widehat{\Gamma} = \max_{\Gamma \in \Omega} \Gamma\,(S_1, S_2). \qquad (6)$$

The performance indices in biometrics authentication system are false acceptance rate denoted FAR which means wrongly identifying an impostor to be an enrollee, and false rejection rate denoted by FRR which means wrongly identifying an enrollee as an imposter.

$$FAR(t) = P(\widehat{w}_1|w_0) = \int_{R_1} p(X|w_0)\,dX = 1 - \int_{R_0} p(X|w_0)\,dX\,, \qquad (7)$$

$$FRR(t) = P(\widehat{w}_0|w_1) = \int_{R_0} p(X|w_1)\,dX \qquad (8)$$

where w_1 denotes the genuine user while w_0 denotes the imposter one. R_0 and R_1 are two exclusive sets in real axis. Both FAR and FRR are desirable to be as low as possible in authentication system. For any biometrics authentication system, whatever classifier takes, there exists a great risk of error. From the viewpoint of Bayesian decision theory, this is represented by the following equations for a two class problem,

$$E(t) = C_r \times FRR(t) + C_a \times FAR(t), \qquad (9)$$

$$E_i(t_i) = C_r^i \times FRR^i(t_i) + C_a^i \times FAR^i(t_i),\ for\ i = 1,..,N \qquad (10)$$

where, N is the total modalities number, C_r denotes the loss function pertinent to the false rejection, and C_a denotes the loss function for the false acceptance. For simplicity, it assume that $C_a = C_a^i\ and\ C_r = C_r^i$.

1) Soft decision level fusion:

The integrated system is denoted by Ψ. The outputs by individual systems Ψ_1 and Ψ_2, are called scores, which stand for the probability of claimant to be a genuine or an imposter. For any fusion strategies, an error is expressed as (9) and (10). If it assumes that $E_1(t_1) \leq E_2(t_2) \leq .. \leq E_N(t_N)$, then it is easily known it is sufficient to prove that $E(t) \leq E_1(t_1)$. For a two-modality and Bayesian rule Fusion:

$$\begin{array}{ll} \text{Decide} & w_0, \text{ if } (X_1, X_2) \in R \\ \text{Decide} & w_1, otherwise \end{array} \qquad (11)$$

where $R = \{\,(X_1,X_2)|C_r p(X_1,X_2|w_0) \geq C_a p(X_1,X_2|w_1)\,\}$.since Ψ_1 and Ψ_2 are independent, it have:

$$p(X_1, X_2|w_0) = p_1(X_1|w_0)p_2(X_2|w_0), \qquad (12)$$

&

$$p(X_1, X_2|w_1) = p_1(X_1|w_1)p_2(X_2|w_1), \qquad (13)$$

Then:

$$\begin{aligned} FAR(t) &= 1 - \int_{R_0} p(X_1, X_2|w_0)dX_1 dX_2 \\ &= 1 - \int_{R_0} p_1(X_1|w_0)dX_1 \int_{R_0} p_2(X_2|w_0)dX_2 \qquad (14) \\ &= 1 - \left(1 - FAR^1(t_1)\right)\left(1 - FAR^2(t_2)\right). \end{aligned}$$

&

$$\begin{aligned} FRR(t) &= 1 - \int_{R_0} p(X_1, X_2|w_1)dX_1 dX_2 \\ &= 1 - \int_{R_0} p_1(X_1|w_1)dX_1 \int_{R_0} p_2(X_2|w_1)dX_2 \qquad (15) \\ &= FRR^1(t_1)FRR^2(t_2). \end{aligned}$$

From the Equations (14) & (15) it can be obviously seen that: $FAR(t) = FAR^1(t_1)$ and $FAR(t) = FAR^2(t_2)$. Thus the two combined modalities cannot improve the false acceptance rate by the Bayesian decision rule. Otherwise $FRR(t) = FRR^1(t_1)$ and $FRR(t) = FRR2t2$. Hence the false rejection rate of the combined system is reduced compared to individual sub-classifiers.

2) Maximum Likelihood Parameter Estimation:

Given a set of observation data in a matrix X and a set of observation parameters θ the ML parameter estimation aims at maximizing the likelihood $L(\theta)$or log likelihood of the observation data $X = \{X_1, \ldots, X_n\}$

$$\hat{\theta} = arg \max_{\theta} L(\theta). \qquad (16)$$

Assuming that it has independent, identically distributed data, it can write the above equations as:

$$L(\theta) = p(X|\theta) = p(X_1, \ldots, X_n|\theta) = \prod_{i=1}^{n} p(X_i|\theta). \qquad (17)$$

The maximum for this function can be find by taking the derivative and set it equal to zero, assuming an analytical function.

$$\frac{\partial}{\partial\theta} L(\theta) = 0. \qquad (18)$$

The incomplete-data log-likelihood of the data for the mixture model is given by:

$$L(\theta) = log(X|\theta) = \sum_{i=1}^{N} log(x_i|\theta) \qquad (19)$$

which is difficult to optimize because it contains the log of the sum. If it considers X as incomplete, however, and posits the existence of unobserved data items $Y = \{y_i\}_{i=1}^{N}$ whose values inform us which component density generated each data item, the likelihood expression is significantly simplified. That is, it assume that $y_i \in \{1\,..K\}$ for each i, and $y_i = k$ if the i-th sample was generated by the k-th mixture component. If it knows the values of Y, it obtains the complete-data log-likelihood, given by:

$$L(\theta, Y) = \log p(X, Y|\theta) \qquad (20)$$

$$= \sum_{i=1}^{N} \log p(x_i, y_i|\theta) \qquad (21)$$

$$= \sum_{i=1}^{N} \log\big(p(y_i|\theta)p(x_i|y_i, \theta)\big) \qquad (22)$$

$$= \sum_{i=1}^{N}\big(\log\ p_{y_i} + \log g\big(x_i|\mu_{y_i}, \Sigma_{y_i}\big)\big) \qquad (23)$$

which, given a particular form of the component densities, can be optimized using a variety of techniques [21].

3) EM algorithm:
The expectation-maximization (EM) algorithm [22][23][24][25] is a procedure for maximum-likelihood (ML) estimation in the cases where a closed form expression for the optimal parameters is hard to obtain. This iterative algorithm guarantees the monotonic increase in the likelihood L when the algorithm is run on the same training database.

The probability density of the Gaussian mixture of k components in κ^d can be described as follows:

$$\Phi(x) = \sum_{i=1}^{N} \pi_i\, \emptyset(x|\theta_i) \quad \forall x \in\ \kappa^d\ , \qquad (24)$$

where $\emptyset(x|\theta_i)$ is a Gaussian probability density with the parameters $\theta_i = (m_i, \sum_i)$, m_i is the mean vector and $\sum_i$ is the covariance matrix which is assumed positive definite given by:

$$\emptyset(x|\theta_i) = \emptyset(x|m_i\,, \Sigma_i) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_i|^{\frac{1}{2}}}\, e^{-\frac{1}{2}(x-m_i)^T\Sigma_i^{-1}(x-m_i)}\,, (25)$$

and $\pi_i \in [0,\ 1]\ (i = 1,2, \ldots, k)$ are the mixing proportions under the constraint $\sum_{i=1}^{k} \pi_i = 1$. If it encapsulate all the parameters into one vector: $\Theta_k = (\pi_1, \pi_2, \ldots, \pi_k, \theta_1, \theta_2, \ldots, \theta_k)$, then , according to Eq. (23), the density of Gaussian mixture can be rewritten as:

$$\Phi(x|\Theta_k) = \sum_{i=1}^{k} \pi_i \emptyset(x|\theta_i) = \sum_{i=1}^{k} \pi_i \emptyset(x|m_i, \Sigma_i). \qquad (26)$$

For the Gaussian mixture modeling, there are many learning algorithms. But the EM algorithm may be the most well-known one. By alternatively implementing the E-step to estimate the probability distribution of the unobservable random variable and the M-step to increase the log-likelihood function, the EM algorithm can finally lead to a local maximum of the log-likelihood function of the model. For the Gaussian mixture model, given a sample data set $S = \{x_1, x_2, \cdots, x_N\}$ as a special incomplete data set, the log-likelihood function can be expressed as follows:

$$\log p(S|\Theta_k) = \log \prod_{t=1}^{N} \emptyset(x_t|\Theta_k) = \sum_{t=1}^{N} \log \sum_{i=1}^{k} \pi_i \emptyset(x_t|\theta_i), \qquad (27)$$

which can be optimized iteratively via the EM algorithm as follows:

$$P(j|x_t) = \frac{\pi_j \emptyset(x_t|\theta_j)}{\sum_{i=1}^{k} \pi_i \emptyset(x_t|\theta_i)}, \qquad (28)$$

$$\pi_j^+ = \frac{1}{N} \sum_{t=1}^{N} P(j|x_t), \qquad (29)$$

$$\mu_j^+ = \frac{1}{\sum_{t=1}^{N} P(j|x_t)} \sum_{t=1}^{N} P(j|x_t) x_t, \qquad (30)$$

$$\Sigma_j^+ = \frac{1}{\sum_{t=1}^{N} P(j|x_t)} \sum_{t=1}^{N} P(j|x_t)\big(x_t - \mu_j^+\big)\big(x_t - \mu_j^+\big)^T. \quad (31)$$

Although the EM algorithm can have some good convergence properties in certain situations, it certainly has no ability to determine the proper number of the components for a sample data set because it is based on the maximization of the likelihood.

4) Greedy EM Algorithm:
The greedy algorithm (GEM) [22][23] [24][26] starts with a single component and then adds components into the mixture one by one. The optimal starting component for a Gaussian mixture is trivially computed, optimal meaning the highest training data likelihood. The algorithm repeats two steps: insert a component into the mixture, and run EM until convergence. Inserting a component that increases the likelihood the most is thought to be an easier problem than initializing a whole near-optimal distribution. Component insertion involves searching for the parameters for only one component at a time. Recall that EM finds a local optimum for the distribution parameters, not necessarily the global optimum which makes it initialization dependent method.

Given p_C a C-component Gaussian mixture with parameters θ_C. the general greedy algorithm for Gaussian mixture is as follows:

1. *Compute the optimal (in the ML sense) one-component mixture* $p_{\mathbf{1}}$ *and set* $C \leftarrow 1$.
2. *Find a new component* $\mathcal{N}(x;\, \mu', \Sigma')$ *and corresponding mixing weight* α' *that increase the likelihood the most:*

$$\{\mu', \Sigma', \alpha'\} = \arg \max_{\{\mu,\Sigma,\alpha\}} \sum_{i=1}^{N} \ln\big[(1-\alpha)p_C(x_i) + \alpha\mathcal{N}\big(x_{i;}\ \mu, \Sigma\big)\big] \qquad (32)$$

while keeping p_C *fixed.*

3. *Set*

   $p_{C+1}(x) \leftarrow (1-\alpha')p_C(x) + \alpha'\mathcal{N}(x;\mu',\Sigma')$ *and then* $C \leftarrow C + 1$.
4. *Update* p_C *using* ***EM*** *(or more other method) until convergence.*
5. *Evaluate some stopping criterion; go to step 2 or quit.*

The stopping criterion in Step 5 can be for example any kind of model selection criterion, wanted number of components, or the minimum message length criterion.

The crucial point is of course Step 2. Finding the optimal new component requires a global search, which is performed by creating CN_{cand} candidate components. The number of candidates will increase linearly with the number of components C, having N_{cand} candidates per each existing component. The candidate resulting in the highest likelihood when inserted into the (previous) mixture is selected. The parameters and weight of the best candidate are then used in Step 3 instead of the truly optimal values.

The candidates for executing Step 2 are initialized as follows: the training data set X is partitioned into C disjoints data sets $\{A_c\}, c = 1 \dots C$, according to the posterior probabilities of individual components; the data set is Bayesian classified by the mixture components. From each A$_c$ number of N_{cand} candidates are initialized by picking uniformly randomly two data points x_l and x_r in A$_c$. The set A$_c$ is then partitioned into two using the smallest distance selection with respect to x_l andx_r. The mean and covariance of these two new subsets are the parameters for two new candidates. The candidate weights are set to half of the weight of the component that produced the set A$_c$. Then new x_l and x_r are drawn until N_{cand} candidates are initialized with A$_c$. The partial **EM** algorithm is then used on each of the candidates. The partial **EM** differs from the **EM** and **CEM** algorithms by optimizing (updating) only one component of a mixture; it does not change any other components. In order to reduce the time complexity of the algorithm a lower bound on the log-likelihood is used instead of the true log-likelihood. The lower-bound log-likelihood is calculated with only the points in the respective set A$_c$. The partial **EM** update equations are as follows:

$$w_{i,C+1} = \frac{\alpha\ \mathcal{N}(x_i,\mu,\Sigma)}{(1-\alpha)p_C(x) + \alpha\ \mathcal{N}(x_i,\mu,\Sigma)} \qquad (33)$$

$$\alpha = \frac{1}{\eta(A_c)} \sum_{i\in A_c} w_{i,C+1} \qquad (34)$$

$$\mu = \frac{\sum_{i\in A_c} w_{i,C+1} x_i}{\sum_{i\in A_c} w_{i,C+1}} \qquad (35)$$

$$\Sigma = \frac{\sum_{i\in A_c} w_{i,C+1}(x_i-\mu)(x_i-\mu)^T}{\sum_{i\in A_c} w_{i,C+1}} \qquad (36)$$

where $\eta(A_c)$ is the number of training samples in the set A$_c$. These equations are much like the basic **EM** update equations in Eqs. (29) - (31). The partial **EM** iterations are stopped when the relative change in log-likelihood of the resulting $C + 1$ –component mixture drops below threshold or maximum number of iterations is reached. When the partial **EM** has converged the candidate is ready to be evaluated.

5) Figueiredo-Jain Algorithm:

The Figueiredo-Jain (FJ) [22[[23][24][26] algorithm tries to overcome three major weaknesses of the basic EM algorithm. The EM algorithm presented previous section requires the user to set the number of components and the number will be fixed during the estimation process. The FJ algorithm adjusts the number of components during estimation by annihilating components that are not supported by the data. This leads to the other EM failure point, the boundary of the parameter space. FJ avoids the boundary when it annihilates components that are becoming singular. FJ also allows starting with an arbitrarily large number of components, which tackles the initialization issue with the EM algorithm. The initial guesses for component means can be distributed into the whole space occupied by training samples, even setting one component for every single training sample.

The classical way to select the number of mixture components is to adopt the "model-class/model" hierarchy, where some candidate models (mixture pdf's) are computed for each model-class (number of components), and then select the "best" model. The idea behind the FJ algorithm is to abandon such hierarchy and to find the "best" overall model directly. Using the minimum message length criterion and applying it to mixture models leads to the objective function:

$$\Lambda(\theta,X) = \frac{V}{2}\sum_{c:\propto_c>0} \ln\left(\frac{N\alpha_c}{12}\right) + \frac{C_{nz}}{2}\ln\frac{N}{12} + \frac{C_{nz}(V+1)}{2} - \ln\mathcal{L}(X,\theta) \qquad (37)$$

Where N is the number of training points, V is the number of free parameters specifying a component, and C_{nz} is the number of components with nonzero weight in the mixture ($\propto_c > 0$). θ in the case of Gaussian mixture is the same as in (Eq. 3) the last term $\ln\mathcal{L}(X,\theta)$ is the log-likelihood of the training data given the distribution parameters (Eq. 23).

The EM algorithm can be used to minimize (Eq. 37) with a fixed C_{nz} . It leads to the M-step with component weight updating formula:

$$\propto_c^{i+1} = \frac{\max\left\{0,\left(\sum_{n=1}^{N} w_{n,c}\right)-\frac{V}{2}\right\}}{\sum_{j=1}^{C}\max\left\{0,\left(\sum_{n=1}^{N} w_{n,c}\right)-\frac{V}{2}\right\}}. \qquad (38)$$

This formula contains an explicit rule of annihilating components by setting their weights to zero.

The above M-steps are not suitable for the basic EM algorithm though. When initial C is high, it can happen that all weights become zero because none of the components have enough support from the data. Therefore a component-wise EM algorithm (CEM) is adopted. CEM updates the components one by one, computing the E-step (updating W) after each component update, where the basic EM updates all components "simultaneously". When a component is annihilated its probability mass is immediately redistributed strengthening the remaining components.

When CEM converges, it is not guaranteed that the minimum of $\Lambda(\theta, X)$ is found, because the annihilation rule (Eq. 38) does not take into account the decrease caused by decreasing C_{nz}. After convergence the component with the smallest weight is removed and the CEM is run again, repeating until $C_{nz} = 1$. Then the estimate with the smallest $\Lambda(\theta, X)$ is chosen. The implementation of the FJ algorithm uses a modified cost function instead of $\Lambda(\theta, X)$.

$$\Lambda'(\theta, X) = \frac{V}{2}\sum_{c:\propto_c>0} \ln \propto_c + \frac{C_{nz}(V+1)}{2} \ln N - \ln \mathcal{L}(X, \theta). \tag{39}$$

IV. EXPERIMENTS AND RESULTS

The experiments were performed using signatures and audio database extracted from video, which is encoded in raw UYVY. AVI 640 x 480, 15.00 fps with uncompressed 16bit PCM audio; mono, 32000 Hz little endian. The capturing devices for recording the video and audio data were: Allied Vision Technologies AVT marlin MF-046C 10 bit ADC, 1/2" (8mm) Progressive scan SONY IT CCD; and Shure SM58 microphone. Frequency response 50 Hz to 15000 Hz. Unidirectional (Cardiod) dynamic vocal microphones. Thirty subjects were used for the experiments in which twenty-six are males and four are females. For each subject, 30 signatures (with dat header) are used. Each line of a (.dat files) consists of four comma separated integer values for the sampled x- and y-position of the pen tip, the pen pressure and the timestamp (in ms); the lines with values of -1 for x, y and pressure represent a pen-up/pen-down event; The device used for recording the handwriting data was a Wacom Graphire3 digitizing tablet. Size of sensing surface is 127.6mm x 92.8mm. With spatial resolution of 2032 lpi (lines per inch), able to measure 512 degrees of pressure. The signature data is acquired with a non-fixed sampling rate of about 100Hz. The audio is extracted as 16 bit PCM WAV file (with wav header), sampled at 16000 Hz, mono little endian. For the audio six multi-lingual (.wav files) of one minute each recording were used for each subject. The database obtained from eNTERFACE 2005 [27]. For signature experts, twenty four signatures from a subject were randomly selected for training, and the other six samples were used for the subsequent validation and testing. Similarly, four samples were used in speech experts for the modeling (training); two samples were used for the subsequent validation and testing. Three sessions of the signature database and speech database were used separately. Session one was used for training the speech and signature experts. Each expert used ten mixture client models. To find the performance, Sessions two and three were used for obtaining expert opinions of known impostor and true claims.

Performance Criteria:

The basic error measure of a verification system is false rejection rate (*FRR*) and false acceptance rate (*FAR*) as defined in the following equations:

False Rejection Rate (***FRR$_i$***): is an average of number of falsely rejected transactions. If ***n*** is a transaction and x(n) is the verification result where 1 is falsely rejected and 0 is accepted and ***N*** is the total number of transactions then the personal False Rejection Rate for user ***i*** is

$$FRR_i = \frac{1}{N}\sum_{n=1}^{N} x(n) \tag{40}$$

False Acceptance rate (***FAR$_i$***) is an average of number of falsely accepted transactions. If ***n*** is a transaction and x(n) is the verification result where 1 is a falsely accepted transaction and 0 is genuinely accepted transaction and ***N*** is the total number of transactions then the personal False Acceptance Rate for user ***i*** is

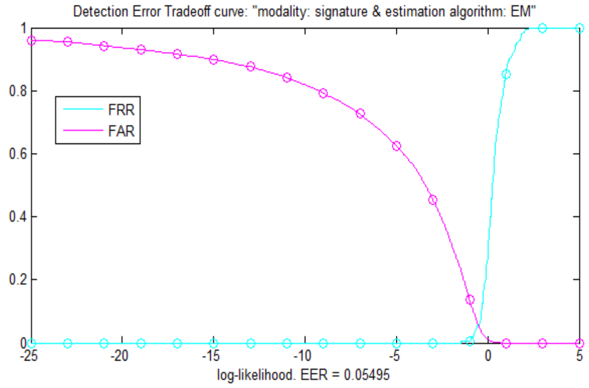
$$FAR_i = \frac{1}{N}\sum_{n=1}^{N} x(n) \tag{41}$$

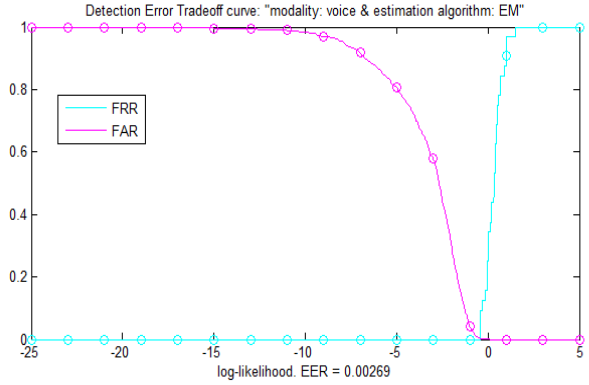
Both FRR_i and FAR_i are usually calculated as averages over an entire population in a test. If P is the size of populations then these averages are

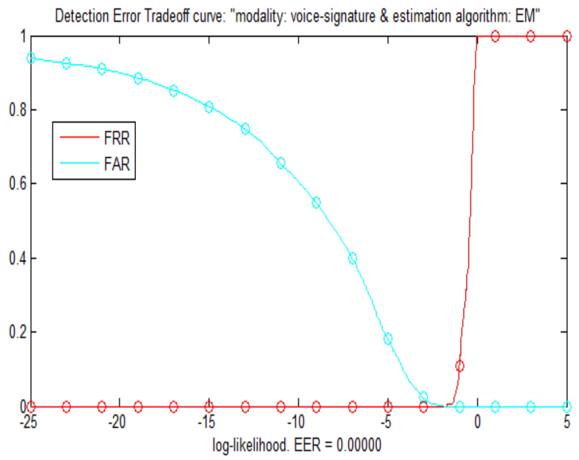
$$FRR = \frac{1}{P}\sum_{i}^{P} FRR_i \tag{42}$$

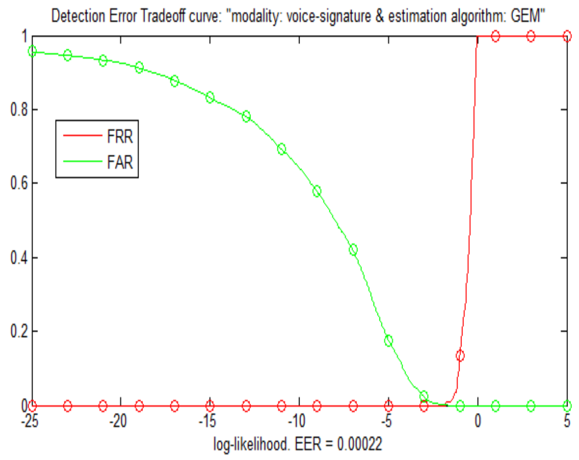
$$FAR = \frac{1}{P}\sum_{i}^{P} FAR_i \tag{43}$$

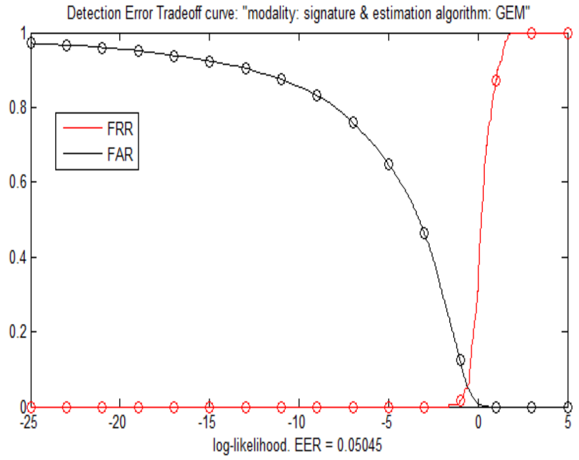
Equal Error Rate (***EER***), is an intersection where FAR and FRR are equal at an optimal threshold value. This threshold value shows where the system performs at its best (see Figure 6).

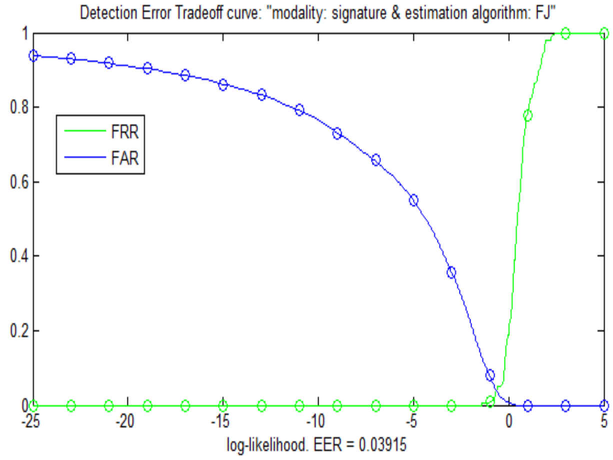


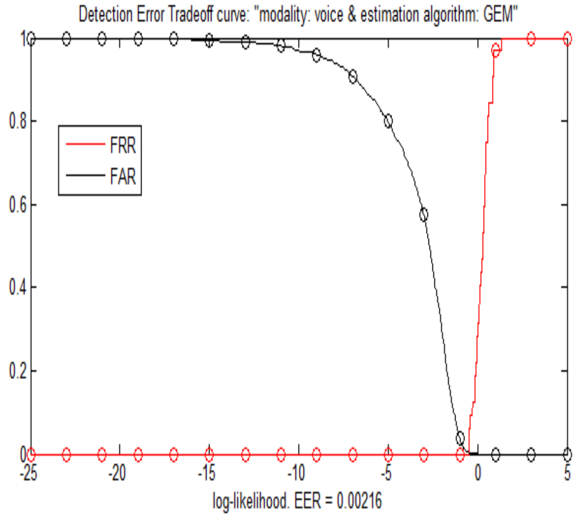


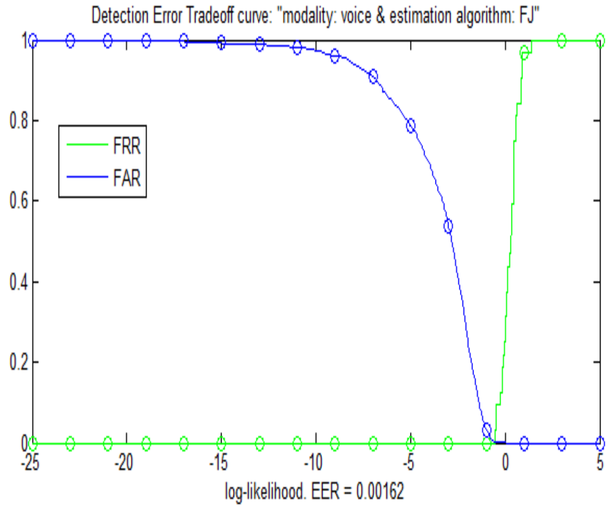

Detection Error Tradeoff curve: "modality: voice-signature & estimation algorithm: EM"
FRR
FAR
1
0.8
0.6
0.4
0.2
0
-25
-20
-15
-10
-5
0
5
log-likelihood. EER = 0.00000


Detection Error Tradeoff curve: "modality: voice-signature & estimation algorithm: GEM"
FRR
FAR
1
0.8
0.6
0.4
0.2
0
-25
-20
-15
-10
-5
0
5
log-likelihood. EER = 0.00022


Detection Error Tradeoff curve: "modality: signature & estimation algorithm: GEM"
FRR
FAR
1
0.8
0.6
0.4
0.2
0
-25
-20
-15
-10
-5
0
5
log-likelihood. EER = 0.05045


Detection Error Tradeoff curve: "modality: signature & estimation algorithm: FJ"
FRR
FAR
1
0.8
0.6
0.4
0.2
0
-25
-20
-15
-10
-5
0
5
log-likelihood. EER = 0.03915


Detection Error Tradeoff curve: "modality: voice & estimation algorithm: GEM"
FRR
FAR
1
0.8
0.6
0.4
0.2
0
-25
-20
-15
-10
-5
0
5
log-likelihood. EER = 0.00216


Detection Error Tradeoff curve: "modality: voice & estimation algorithm: FJ"
FRR
FAR
1
0.8
0.6
0.4
0.2
0
-25
-20
-15
-10
-5
0
5
log-likelihood. EER = 0.00162

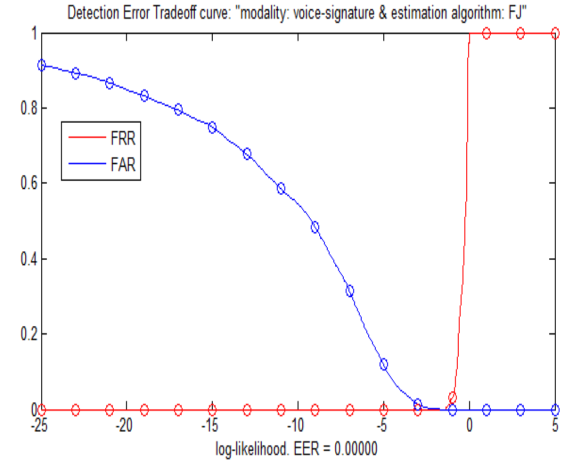


Figure 6. Detection error tradeoff curves (DET).

As a common starting point, classifier parameters were selected to obtain performance as close as possible to *EER* on clean test data (following the standard practice in the Biometrics verification area of using *EER* as a measure of expected performance). A good decision is to choose the decision threshold such as the false accept equal to the false reject rate. In this paper it uses the Detection Error Tradeoff (DET) curve to visualize and compare the performance of the system.

V. CONCLUSION

The paper has presented a human authentication method combined behavioural signature and speech information in order to improve the problem of single biometric authentication, since single biometric authentication has the fundamental problems of high FAR and FRR. It has presented a framework for fusion of match scores in multi-modal biometric system based on soft decision level fusion. The (EM), (GEM) and (FJ) estimation algorithms achieve a significant performance rates, EER=0.0 % for "EM" and "FJ", EER=0.02 % for "GEM", for the combined modalities. Based on the experimental results, it has shown that EER can be reduced down significantly between the single mode and a combined mode. Thus, the combined behavioral information scheme is more robust and have a discriminating power, which can be explored for identity authentication.

REFERENCES

[1] Sofia Gleni & Panagiotis Petratos, *"DNA Smart Card for Financial Transactions"* The ACM Student Magazine 2004, http://www.acm.org

[2]Girija Chetty and Michael Wagner, *"Audio-Visual Multimodal Fusion for Biometric Person Authentication and Liveness Verification"*, Copyright © 2006, Australian Computer Society, Inc. This paper appeared at the *NICTA-HCSNet Multimodal UserInteraction Workshop (MMUI2005)*, Sydney, Australia.

[3] Norman Poh and Samy Bengio, *"Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometric Authentication"*, IDIAP RR 04-44, August 2004, a IDIAP, CP 592, 1920 Martigny, Switzerland.

[4] M. Fuentes*, D. Mostefa**, J. Kharroubi**, S. Garcia-Salicetti*, B. Dorizzi*, G. Chollet**; *"IDENTITY VERIFICATION BY FUSION OF BIOMETRIC DATA: ON-LINE SIGNATURES AND SPEECH"*, * INT, dépt EPH, 9 rue Charles Fourier, 91011 EVRY France; **ENST, Lab. CNRS-LTCI, 46 rue Barrault, 75634 Paris – The Advent of Biometrics on the Internet, A COST 275 WORKSHOP. Rome, Italy November 7-8, 2002.

[5] A. Perez-Hernandez, A. Sanchez and J. F. Velez, *"Simplified Stroke-based Approach for Off-line Signature recognition"*, Departamento de Informatica, Estadistica y Telematica - Universdad Rey Juan Carlos, Campus de Mostoles 28933 Mostoles (Madrid),SPAIN. 2nd COST 275 Workshop – Biometrics on the Internet, Vigo, 25-26 March 2004. SPAIN

[6] Hugo Gamboa[a] and Ana Fred[b], *"A Behavioural Biometric System Based on Human Computer Interaction"*, [a]Escola Superior de Tecnologia de Setubal, Campo do IPS, Estefanilha, 2914-508 Setubal, Portugal; [b]Instituto de Telecomunica¸c˜oes, Instituto Superior T´ecnico IST - Torre Norte, Piso 10, Av. Rovisco Pais 1049-001, Lisboa, Portugal

[7] Ibrahim S. I. ABUHAIBA, *"Offline Signature Verification Using Graph Matching"*, Department of Electrical and Computer Engineering, Islamic University of Gaza, P. O. Box 1276, Gaza-PALESTINE

[8] Souheil Ben-Yacoub, Yousri Abdeljaoued, and Eddy Mayoraz, *Member, IEEE, "Fusion of Face and Speech Data for Person Identity Verification",* IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 10, NO. 5, SEPTEMBER 1999

[9] H. J. Korves, L. D. Nadel, B. T. Ulery, D. M. Bevilacqua Masi, "Multi-biometric Fusion: From Research to Operations", *MTS MitreTek Systems*, sigma summar 2005, pp. 39-48, http://www.mitretek.org/home.nsf/Publications/SigmaSummer2005

[10] A. Kounoudes[1], N. Tsapatsoulis[2], Z. Theodosiou[1], and M. Milis[1], *"POLYBIO: Multimodal Biometric Data Acquisition Platform and Security System"*, [1] SignalGeneriX Ltd, Arch.Leontiou A' Maximos Court B', 3rd floor, P.O.Box 51341, 3504, Limassol, Cyprus & [2] Cyprus University of Technology, Arch.Kyprianos Kyprianos, P.O.Box 50329, 3603, Limmasol, Cyprus

[11] Corradini (1), M. Mehta (1), N.O. Bernsen (1), J. C. Martin (2,3), S.Abrilian (2), *"MULTIMODAL INPUT FUSION IN HUMAN-COMPUTER INTERACTION"*, On the Example of the NICE Project 2003. (1) Natural Interactive Systems Laboratory (NISLab), University of Southern Denmark, DK-Odense M, Denmark. (2) Laboratory of Computer Science for Mechanical and Engineering Sciences, LIMSI-CNRS, F-91403 Orsay, France. (3) Montreuil Computer Science Institute (LINC-IUT), University Paris 8, F-93100 Montreuil, France.

[12] Conrad Sanderson, Samy Bengio, Herve Bourlard, Johnny Mariéthoz, Ronan Collobert, Mohamed F. BenZeghiba, Fabien Cardinaux, and S´ebastien Marcel, *"SPEECH & FACE BASED BIOMETRIC AUTHENTICATION AT IDIAP"*, Dalle Molle Institute for Perceptual Artificial Intelligence (IDIAP). Rue du Simplon 4, CH-1920 Martigny, Switzerland.

[13] Claus Vielhauer[a], Sascha Schimke[a], Valsamakis Thanassis[b] ,Yannis Stylianou[b] , [a] Otto-von-Guericke University Magdeburg, Universitaetsplatz 2, D-39106, Magdeburg, Germany, [b] University of Crete, Department of Computer Science, Heraklion, Crete, Greece, *"Fusion Strategies for Speech and Handwriting Modalities in HCI"*, Multimedia on Mobile Devices, edited by Reiner Creutzburg, Jarmo H. Takala, Proc. of SPIE-IS&T Electronic Imaging, Vol. 5684 © 2005

[14] Lasse L. Mølgaard and Kasper W. Jørgensen,*"Speaker Recognition: Special Course"*, IMM_DTU December 14, 2005

[15] S. Davis and P. Mermelstein. *"Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences". IEEE Transactions on Acoustics, Speech and Signal Processing*, (4):357–366, 1980.

[16] Pekka Paalanen, *" Bayesian classification using gaussian mixcute model and EM estimation: implementation and comparisons"*,Information Technology Project, 2004, http://www.it.lut.fi/project/gmmbayes/

[17] D.A. Reynolds, *"Experimental Evaluation of Features for Robust Speaker Identification", IEEE Trans. Speech and Audio Processing*2 (4), 1994, 639-643.

[18] J. Richiardi*, J. Fierrez-Aguilar**, J. Ortiga-Garcia** and A. Drygajlo*, *"On-line signature verification resilience to packet loss in IP networks"*, second COST 275 WORKSHOP Biometrics on the Internet: Fundamentals, Advances and Applications. University of Vigo, Vigo-Spain 25-26 March 2004. *Swiss Fderal Institute of Technology of Lausanne, Switzerland, **Universidadd Politéctica de Madrid, Spain

[19] Kalyan Veeramachaneni, Lisa Ann Osadciw, and Pramod K. Varshney, *"An Adaptive Multimodal Biometric Management Algorithm"*, IEEE TRANSACTIONS ON SYSTEMS, MAN,

AND CYBERNETICS-PART C: APPLICATIONS AND REVIEWS, VOL. 35, NO. 3, AUGUST 2005

[20] Dongliang Huang, Henry Leung and Winston Li, *" Fusion of Dependent and Independent Biometric Information Sources",* Department of Electrical & Computer Engineering - University of Calgary, 2500 University Dr NW Calgary, AB T2N 1N4

[21] Kittler, J., Hatef, M., Duin, R. P. W. and Matas, J., *"On combining classifiers"*. IEEETransactions on Pattern Analysis and Machine Intelligence, 20(3): 226–239. 1998

[22] Kalyan Veeramachaneni, Lisa Ann Osadciw, and Pramod K. Varshney, *"An Adaptive Multimodal Biometric Management Algorithm"*, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART C: APPLICATIONS AND REVIEWS, VOL. 35, NO. 3, AUGUST 2005

[23] Van Trees, Harry L., *"Detection, Estimation, and Modulation Theory*", Part I, John Wiley and Sons, 1968.

[24] Qing Yan and Rick S. Blum, "Distributed Signal Detection under the Neyman-Pearson Criterion" , EECS Department Lehigh University Bethlehem, PA 18015

[25] P. Paalanen, J.-K. Kamarainen, J. Ilonen, H. Kälviäinen, *"Feature Representation and Discrimination Based on Gaussian Mixture Model Probability Densities: Practices and Algorithms",* Department of Information Technology, Lappeenranta University of Technology, P.O.Box 20, FI-53851 Lappeenranta, Finland 2005

[26] S. Davis and P. Mermelstein. "*Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*". *IEEE Transactions on Acoustics, Speech and Signal Processing*, (4):357–366, 1980.

[27] Yannis Stylianou, Yannis Pantazis, Felipe Calderero, Pedro Larroy, Francois Severin, Sascha Schimke, Rolando Bonal, Federico Matta, and Athanasios Valsamakis, *"GMM-Based Multimodal Biometric Verification"*, eNTERFACE 2005 The summer Workshop on Multimodal Interfaces July 18th – August 12th, Facultè Polytechnique de Mons, Belgium.