Analyzing a multimodal biometric system using real and virtual users

Tobias Scheidat, Claus Vielhauer
Dept. of Computer Science, Univ. of Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany

ABSTRACT

Three main topics of recent research on multimodal biometric systems are addressed in this article: The lack of sufficiently large multimodal test data sets, the influence of cultural aspects and data protection issues of multimodal biometric data. In this contribution, different possibilities are presented to extend multimodal databases by generating so-called virtual users, which are created by combining single biometric modality data of different users. Comparative tests on databases containing real and virtual users based on a multimodal system using handwriting and speech are presented, to study to which degree the use of virtual multimodal databases allows conclusions with respect to recognition accuracy in comparison to real multimodal data. All tests have been carried out on databases created from donations from three different nationality groups. This allows to review the experimental results both in general and in context of cultural origin. The results show that in most cases the usage of virtual persons leads to lower accuracy than the usage of real users in terms of the measurement applied: the Equal Error Rate. Finally, this article will address the general question how the concept of virtual users may influence the data protection requirements for multimodal evaluation databases in the future.

Keywords: biometrics, data protection, multimodal, handwriting, speech, real user, virtual user, verification

1. INTRODUCTION

Biometric authentication systems provide an alternative to the conventional authentication methods, secret knowledge or personal possession. The fact that the authentication object is directly linked with the person itself (*passive biometrics*: e.g. fingerprint, face) or with the behavior of the person (*active biometrics*: e.g. signature, voice) is one main advantage of biometrics. Theft or handoff (intended or accidental) of biometric authentication objects are not possible in an easy way, on the other side these are central problems of using secret knowledge or personal possession. An idea to solve these problems is to combine at least two of the three authentication factors mentioned above, such combination of a personal identification number (secret knowledge) and a smart card (personal possession) for example.

In general a biometric system works in two operation modes: enrollment or authentication. The enrollment means the registration of a user within the system where the reference data are stored for the user and associated with the user's identity. Figure 1 shows the data acquisition module captures the physiological or behavioral trait of the user and after an optional preprocessing the feature extraction module determines a feature set from the captured data describing the current biometric data within the biometric system. Then this feature set is stored as reference data in the database of the biometric system. For authentication the same process steps are carried out up to the feature extraction. The matching module compares the feature vectors of authentication data and reference data, and calculates a similarity value, the so-called matching score. This score is the basis for the decision whether the user is the person which he or she claims to be (*verification mode*), or who the user is (*identification mode*).

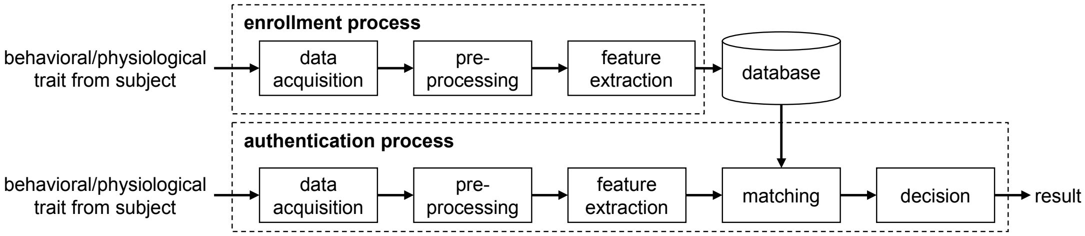


Figure 1. Scheme of a general biometric authentication process

Security, Steganography, and Watermarking of Multimedia Contents IX,
edited by Edward J. Delp III, Ping Wah Wong, Proc. of SPIE-IS&T Electronic Imaging,
SPIE Vol. 6505, 650512, © 2007 SPIE-IS&T · 0277-786X/07/$18

In recent research the interest on multimodal biometric systems for automatic user authentication rose strongly. The aim of combination of different biometric systems is the possible improvement of the authentication performance in comparison to the best single system involved. Other goals addressed by multimodal biometric systems are to make spoof attacks difficult for an imposter or to provide one (or more) alternative modality if a trait is missing or a trait can be recognized poorly. For multimodal biometrics there are three levels where fusion can be carried out based on the point of fusion within the biometric authentication process: fusion on feature extraction level, fusion on matching score level and fusion on decision level. The fusion on feature extraction level is based on a combination of the single feature vectors of the different systems, by concatenation of the vectors for example. However, this is an unpopular method because of the high dimension of the joint feature vector and the consequently high effort in calculation of the matching score. For the fusion on matching score level the scores of the systems involved are combined to one single joint matching score. Most multimodal fusions use this point within the authentication process for fusion because of its advantages such as one individual match score of each subsystem, simple possible weighting strategies based on this scores or one single value as basis for the decision step. The entire authentication process of each system is carried out at the fusion on decision level, and is based on the single decisions one common result is determined, e.g. by Boolean operations. Jain and Ross describe in [1] an improvement by a multimodal fusion using face, fingerprint and hand geometry. In [2] Vielhauer et al. present a multimodal system where a speech recognition system and a signature recognition system are fused in order to obtain a better authentication result in comparison to the single biometric systems involved. An enhancement of this multimodal system was suggested in [3] by exchange of the single signature component by a multi-algorithmic handwriting subsystem. By this multimodal/multi-algorithmic fusion an improvement of *15%* could achieved in comparison to original multimodal system described in [2]. The multi-algorithmic method is proposed in [4] by Scheidat et al. and uses a combination of four signature verification algorithms in order to improve the verification result. The best fusion strategy results in a decrease of the performance measure, the equal error rate, of *12.1%* in comparison to the best individual algorithm.

At the evaluation of multimodal systems recurrently arise the problem of obtaining suitable test data of a single person for the used modalities. Existing multimodal biometric databases which contain the required modalities in the required quality and quantity are rare. Therefore Wolf et al. propose in [5] a possibility for a multimodal database's enhancement by building virtual users and using these to enlarge the database. Virtual users are considered as the combination of two or more traits captured from different users. The article shows that such an enlargement by approximately *50%* leads to degradations of up to approximately *25%* with respect to equal error rate based on a multimodal fusion of handwriting and speech. In this paper we propose the creation of an entire database of virtual users by shuffling the collected data of handwriting and speech. The underlying system mixes the existing data of the individual persons without the data of a single person being combined with each other.

Recently, another area of research is the cross-cultural evaluation of active biometrics. The general idea in this domain is to include additional non-biometric information about users of biometric systems, such as cultural background (e.g. spoken and written language, nationality), biological and physiological data (e.g. gender, handedness), as well as technical characteristics of the system itself (e.g. sensor type) in the biometric processes. A methodology for this purpose has been suggested based on a structured set of metadata ([6]) and based on experimental evaluations. The authors show that the recognition accuracy of a biometric handwriting recognition system may vary significantly, depending on metadata such as gender or written language. In [7] Jain et al. describe the utilization of "soft" biometric traits like gender or height to add the identity information to the primary biometric like fingerprint, face or hand-geometry. But here no inter-cultural context is proposed, however the improvement of significantly in general. Wolf et al. show in [8] that more discrimination parameters can be found analyzing the cross-cultural impact on behavioral biometric data. Also parameters like changing experiences and learnt attributes of writing show distinct influences. The field of inter-cultural and multi-modal user interfaces was extended by the aspect of metadata. In [9] Scheidat et al. show furthermore, that it is possible to give distinct design recommendations for active handwriting biometrics according to user groups of specific cultural background with respect to different security levels including forgery scenarios.

Besides the construction of multimodal databases with the desired modalities based on virtual users and/or expansion of an existing database, the aspect of the protection of data privacy arose recently. In detail building virtual users may enhance data protection of personal data and therefore could enable research without legal restriction. In Europe the basic of current data protection laws was set by the Directive 95/46 on the Protection of Individuals with regard to the Processing of Personal Data of the European Union ([10]). There are different points of view in the classification of data protection requirements concerning biometric data of individuals. One possible interpretation of German law is that data can be differentiated in related and relatable data of an individual. For example, from a picture of a person's face one can typically directly recognize and identify the person if he or she is known. Consequently this picture contains related data of an individual. On the other hand the solely information of a fingerprint image can not be used easily for identification without help and/or information from a third person and/or a technical authentication system. These data are described as relatable data of a person. From the point of view of German law needs related data a higher degree of protection than relatable data. The differentiation between relatable data and related data is blurry: If in a multimodal database the picture of the face of a person is stored in addition to the fingerprint image of the same person with a direct relation between them, the entire data set of this individual is related data and the need of its protection rises strongly.

In this article we study the possibility of creation of virtual users in order to avoid the cross relation between the biometric data with the class of the test subjects in a biometric multimodal database. By avoiding these cross relations the necessary effort of data protection may decrease. In addition, a database of virtual users can be formed by mixing unimodal data from different unimodal and/or multimodal databases. We analyze the authentication performance of such virtual users' database using biometric handwriting and speech data. Real and virtual multimodal databases are compared in order to find out if it is an alternative to a multimodal database holding data only of real users.

This paper is structured as follows: In the second section the handwriting and the speech based systems are described shortly, and fundamentals and strategies of biometric fusion are given. Section three describes the evaluation methodology, the underlying multimodal database and biometric error rates as performance measure for the evaluation. The test results and a discussion of their meaning are shown in section four. A short summary of this paper and an outlook of future work are given in section five.

2. MULTIMODAL BIOMETRIC FUSION

In this section the fundamentals of fusion of the handwriting and speech subsystems are described. Firstly, a general operation breakdown is given of the underlying algorithms, Biometric Hash for handwriting recognition and Mel-Frequency Cepstrum Coefficients for speech recognition. Secondly, the weighted fusion strategy for combing both subsystems is presented.

2.1 Verification algorithms

The verification algorithm for the online handwriting modality is based on the Biometric Hash algorithm, as introduced in detail in [11] and [12]. In general, this method determines a statistical feature vector of *k=69* statistical parameters (online and offline features), which are transformed into the hash value space by an interval mapping function. This mapping, denoted as Key Generation, results in a feature vector representation $\vec{b}=(b_1,...,b_k)$ supported by a user specific statistical model, consisting of an Interval Matrix (*IM*) and a Masking Vector (*mv*), which is obtained during enrollment.

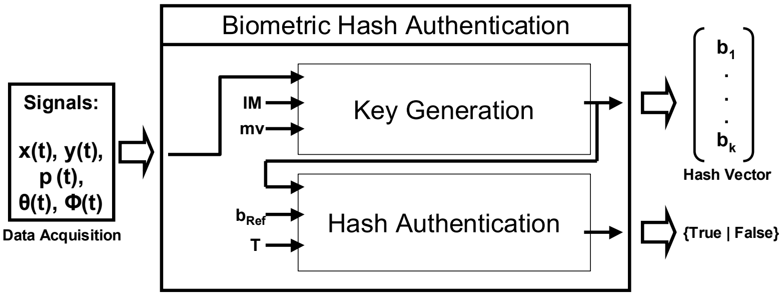


Figure 2. User Authentication based on Biometric Hash

As shown in the left part of figure 2, during verification five discrete signals based on measurements of horizontal and vertical pen position *x(t)* and *y(t)*, pen tip pressure *p(t)* and pen azimuth and altitude *Θ(t)* and *Φ(t)* respectively are taken from the digitizer tablet. Based on these five signals, the Key Generation module will calculate an actual feature vector $\vec{b}$, which is compared to a stored reference vector $\vec{b}_{\mathrm{Ref}}$ against some decision threshold value T in the Hash

Authentication Module. In our system, this authentication is performed by calculation of the Hamming Distance between the two vectors. Finally this verification method results in a binary True/False decision with respect to the actual biometric data and the given threshold.

The speaker verification uses Mel-Frequency Cepstrum Coefficients (MFCC), aspiring to model sounds. As it is known the higher the frequency of two sounds the more difficult it becomes to distinguish them, a logarithmic scale – the mel scale [13] – for the perceived pitch is used. Also rather than just using the spectrum of the signal the spectrum of the log spectrum – the cepstrum [14] – of the signal is used.
The method first separates the signal at every 10ms in frames of 30ms length with a hamming windowing function:

$$h(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N}, 0 \le n < N \tag{1}$$

Afterwards on the spectrum of frames with an energy exceeding a defined threshold a filter bank was applied. This filter bank consisted of 20 uniformly distributed triangular band pass filters in steps of approximately 135.2 mels. Of this mel-frequency wrapped spectrum Ψ the MFCC was calculated as follows:

$$MFCC = \sum_{l=1}^{20} \log \Psi(l) \cos\left[\frac{(l+0,5)k\pi}{20}\right], 0 \le k < 20 \tag{2}$$

An enrollment sample will represented as a set of 32 centroids retrieved from the MFCC vector set with the LBG algorithm presented by Linde et al. in [15]. The score between a verification/attack sample and an enrollment sample will be minimum squared euclidean distance between each of the verification sample's MFCCs and each of 32 enrollment centroids.

2.2 Fusion strategies

The fusion of handwriting and speech is carried out after the matching score computation within the verification process (fusion on matching score level). An important advantage lies here in the possibility of weighting the individual matching scores derived from each subsystem. For the fusion one of five weighting strategies presented in previous work ([4]) is used for multi-algorithmic fusion: the linear weighted fusion. With this strategy the subsystems are weighted by the relations of their empirical determined equal error rates (EERs). The EER is an evaluation parameter which is generally used for comparison of the authentication performance of biometric systems and is described in more detail in section *3.3 Evaluation Methodology*. At the linear weighted fusion strategy the system, which received the highest EER, gets the smallest weight and contrary. The individual weights are determined according to the following formula:

$$\begin{array}{ll} \textit{Match Scores}: & s_1, s_2, ..., s_n \\ \textit{Weights}: & w_1, w_2, ..., w_n \\ & n = \textit{number of systems involved} \end{array}$$

$$w_i = \frac{eer_i}{\sum_{m=1}^{n} eer_m} \tag{3}$$

$$\begin{array}{ll} \textit{Conditions}: & w_1 + w_2 + ... + w_n = 1 \\ \textit{Fusion}: & s_{fus} = w_1 s_1 + w_2 s_2 + ... + w_{n-1} s_{n-1} + w_n s_n \end{array}$$

In this article we will focus on a limit of *n = 2* modalities, handwriting and speech. The joint matching score of the weighted fusion is used by the decision module to determine the final authentication result of the whole system by a threshold based comparison.

3. METHODOLOGY

This section describes the basics of the evaluation of the multimodal system: In the first subsection an overview of the used multimodal database is given. The next subsection describes the terms of real and virtual users in more detail and proposes three methods to create virtual multimodal users based on single and/or multimodal biometric modality data. The biometric error rates used as authentication performance measure in this article are described in the third subsection.

3.1 Multimodal database

In this study an existing database storing multimodal biometric data (handwriting and speech) was used for researching the impacts of virtual users on verification algorithms and their fusion. The data used in this work was captured in three different countries (Germany, India and Italy) in English language within the research project CultureTech ([16]). Along with the biometric data, metadata describing the personal background of the data subjects and the technical environment was also captured, linked with the user's identity and finally stored in the database. As shown before in [8] the database stores samples of different semantics donated as speech and handwriting. The metadata *semantics* here are alternative written or spoken contents like predefined PINs, given sentences or signature for handwriting and good name for speech. The semantics are based on different tasks, such as to give individual answers to questions, to write or speak a given sentence, word or number, or to draw an individual symbol. In CultureTech multimodal database the number of written semantics amounts *49*, and for the speech modality there are *46* semantics. Figure 3 shows the scheme of acquiring and processing handwriting and speech data for enrollment and verification. During the enrollment's data acquisition the metadata concerning personal and technical background are also determined and stored in the database. Later the metadata can be used to generate test sets according special evaluation scenarios, using a combination of one semantic of one nationality group for example.

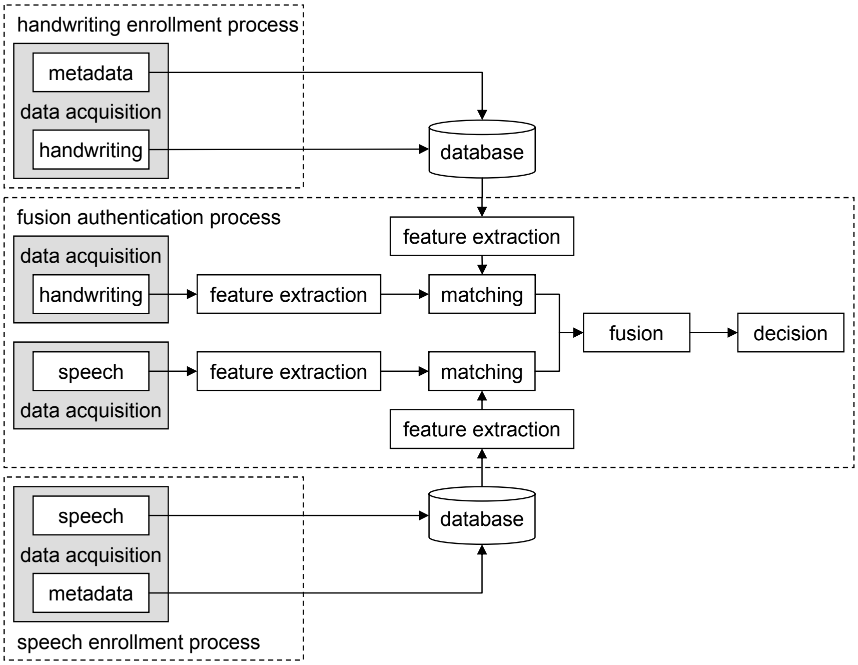
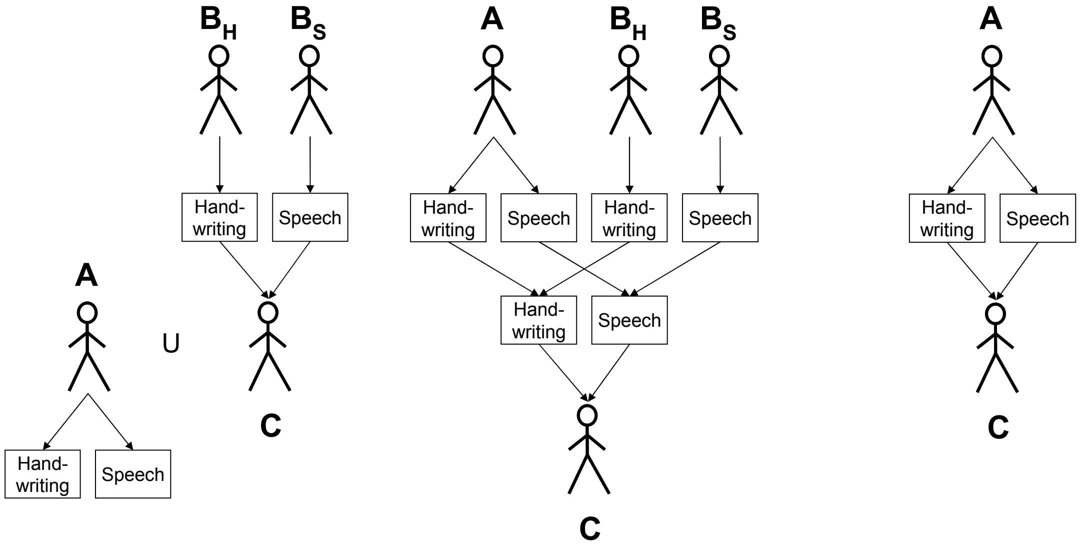


Figure 3. Scheme of enrollment and authentication process of multimodal system used for evaluation

In order to examine the influence of virtual users on the authentication result, three groups are formed consisting of the participants of the individual nationalities: German, Indian and Italian. A fourth group is the union set of all test persons without consideration of their nationality. Further three semantics were selected from the biometric data of these three groups in order to study the impact of different semantics and nationalities. The semantic *signature* describes the signature recorded in handwriting modality and the good name recorded in speech modality and represent individual semantics that differ from test subject to test subject. A predefined *PIN* is given as "7-79-93" in both handwriting and speech, as well as the given *sentence* "Hello, how are you?".

3.2 Real users and virtual users

As shown in figure 4 the modalities of handwriting and speech one can differentiate between three types of users: A user of type *A* donated both, handwriting and speech, for each semantic. Such a type *A* user is a so called *real user*. The users of the second type have donated only one of the two modalities (type *B*), handwriting (*H*) or speech (*S*). Type *C* users are *virtual users* which are built using data from users of type *A* and/or *B*.



(i.) (ii.) (iii.)

Figure 4. Three methodologies of combining biometric data to create virtual users

Figure 4 shows three possibilities to form virtual users from users of the types *A* and/or *B*:

(i.) The data of the users of type *B* are combined by linking handwriting and speech data from the different users to virtual users *C*. In order to enlarge an existing real multimodal database the virtual users can be added to the real users (see figure 4 (i.)). The possible total number of virtual users *C* (*#C*) is the minimum number of the type *B* users which have donated only handwriting ($\#B_H$) or only speech ($\#B_S$), respectively:

$$\# C = Min\ (\# B_{\boldsymbol{H}},\ \# B_{\boldsymbol{S}}) \qquad (4)$$

(ii.) In order to create virtual users (type *C*) the data of the two modalities are combined using the biometric data of the users of the types *A* and *B* under the condition that the originators of both are different persons (figure 4 (ii.)). This method breaks the cross relation between the data of the real users (*A*) and also expand the size of the database. Here the total number of virtual users *C* (*#C*) is determined by the minimum of all users which have given handwriting ($\#A_H$ + $\#B_H$) and/or speech ($\#A_S$ + $\#B_S$) biometrics:

$$\# C = Min((\# A_{\boldsymbol{H}} + \# B_{\boldsymbol{H}}),\ (\# A_{\boldsymbol{S}} + \# B_{\boldsymbol{S}})) \qquad (5)$$

(iii.)The third method uses only the multimodal data of the real users (*A*) to create virtual users (*C*) by recombining the handwriting and speech data (see figure 4 (iii.)). Here only the cross relations of the individuals are destroyed and the number of virtual users *C* (*#C*) is equal to the number of real users *A* (*#A*) as described in equitation (6):

$$\# C = \# A = \# A_H = \# A_S \qquad (6)$$

An evaluation of the scenario described in (i) is presented in [5] by Wolf et al. In this article the evaluation is carried out on virtual users created by method (ii). Possibly (iii) could be the basis for further experiments after finishing this work to study the impact of simple shuffling multimodal data in to generate virtual user database.

As shown in table 1 *24* test sets were created based on *real users* and *virtual users*, metadata *semantic* and metadata *nationality*. In section 4 the results are presented regarding these test sets using biometric error rates as described in subsection 3.3. The data set size of the real users of semantic signature for all test subjects amounts *27*, and after creation of virtual users the number of individuals reaches *40*. For the semantic PIN the real users' database holds *19* persons and the virtual users' database holds *31*. There are *22* real users and *38* virtual users for semantic sentence. As one can see in table 1, the enhancement of the multimodal database by creating virtual users is *32.5%* for signature, *38.71%* for PIN, and *42.11%* for sentence. In general the enlargement of the multimodal database lies between *13%* and *73%* accordingly to metadata semantic and nationality.

Table 1. Number of users divided by modality, real users and virtual users, and relative database enhancement

Semantic	Handwriting	Speech	Real User	Virtual User	Enhancement
German Donors					
Signature	27	18	8	18	55.56 %
PIN	30	14	5	14	64.29 %
Sentence	11	16	3	11	72.73 %
Indian Donors					
Signature	19	15	13	15	13.33 %
PIN	19	10	8	10	20.00 %
Sentence	19	15	13	15	13.33 %
Italian Donors					
Signature	16	7	6	7	14.29 %
PIN	16	7	6	7	14.29 %
Sentence	16	7	6	7	14.29 %
Joint Set of Donors					
Signature	62	40	27	40	32.50 %
PIN	65	31	19	31	38.71 %
Sentence	46	38	22	38	42.11 %

3.3 Evaluation methodology

For our tests we use biometric error rates, where the False Rejection Rate (FRR) indicates how frequently authentic persons are rejected from the system whereas the acceptance rate of non-authentic subjects is represented by the False Acceptance Rate (FAR). The previous mentioned Equal Error Rate (EER) denotes the point of intersection of FRR and FAR characteristics where both error rates yields identical values. The EER is used to compare the results of different test scenarios. Additionally for the evaluation we determine the weights for handwriting and speech based on the EERs of the individual verification results of the single modalities regarding the formula (3) in section 2.2. Using these weights and the matching scores we fuse both systems and can calculate an EER of the multimodal system.

For each user *5* enrollments of each modality are used, holding *4* handwriting or speech samples each. For verification issues *5* additional samples were captured. In the verification mode of the underlying evaluation system the enrollments of one user are compared with each verification sample of the same user. Based on these operations the FRR is captured for each test set. The considering FAR is determined by random attacks. Here each enrollment of one user is compared to each verification sample of all users except the current user. The method described simulates a closed scenario where only persons registered within the biometric system are considered.

4. EVALUATION RESULTS

In this section the evaluation results are presented and discussed based on comparison of real users and virtual users, and metadata semantic and nationality. Because of the limited number of real users as well as virtual users (see table 1) this study is not statistical representative but it shows the functional concept of creating virtual multimodal users based on single and/or multimodal data and evaluating such a virtual database based on additional data, here metadata semantic and nationality.

The tables 2 to 5 show the verification results in respect to the authentication performance measure used, the equal error rate (EER), for real and virtual users. The columns EER_H and EER_S for both, real users and virtual users show the EERs for the single biometric systems using handwriting (*H*) and speech (*S*), respectively. The columns titled *fusion* contain the individual weights ($weight_H$, $weight_S$) of the modalities involved and the *EER* of the fused modalities. The rows hold the results of the semantics *signature*, *PIN* and *sentence*. The handwriting based system always determines the best results in comparison to the speech based system in respect to the EER in all test setups.

In table 2 the results of the German participants are presented. Here the real users' results of the fusion are better than the results of the single modalities for each semantic. The relative improvement by the fusion in comparison to the best single subsystem amounts *0%* for signature, *12.1%* for PIN and *15.7%* for sentence. If one compares the single results of the individual systems for real users and virtual users, one can observe that there are aggravations for handwriting in one out of three cases and for speech in two out of three cases. For example, for the written sentence the EER based on the virtual users' database improves by *154.4%* in comparison to the real users' database. On the other side the EER of spoken sentence degrades by *21.3%* for virtual users compared with real users, however the virtual users based fusion leads to a relative improvement of *138.7%* with an ERR of *0.0222*.

Table 2. Weights and EERs of German real and virtual users (H=handwriting subsystem, S=speech subsystem)

	Real users					Virtual users				
	Single system		Fusion			Single system		Fusion		
Semantic	**EER_H**	**EER_S**	**$weight_H$**	**$weigths_S$**	**EER**	**EER_H**	**EER_S**	**$weight_H$**	**$weigths_S$**	**EER**
Signature	**0.0350**	0.3000	0.896	0.104	**0.0350**	0.0234	0.2737	0.921	0.079	**0.0222**
PIN	0.0648	0.3000	0.882	0.118	**0.0578**	0.0722	0.3479	0.828	0.172	**0.0680**
Sentence	0.0613	0.1466	0.705	0.295	**0.0530**	0.0241	0.1863	0.885	0.115	**0.0222**

The results of the Indian donors are shown in table 3. A general observation, the individual results of the subsystems as well as of the fusion for virtual users and all three semantics are worse than for real users. Another fact is that the fusion using real users' data leads to an improvement in two out of three cases: for signature it degrades by *8.8%*, for PIN and sentence it improves by *7.2%* and *73.3%*, respectively. Based on the virtual users the fusion improves the best result of the systems involved for all three semantics by *10.8%* for signature, *23.3%* for PIN and *20.3%* for sentence.

Table 3. Weights and EERs of Indian real and virtual users (H=handwriting subsystem, S=speech subsystem)

	Real users					Virtual users				
	Single system		Fusion			Single system		Fusion		
Semantic	**EER_H**	**EER_S**	**$weight_H$**	**$weigths_S$**	**EER**	**EER_H**	**EER_S**	**$weight_H$**	**$weigths_S$**	**EER**
Signature	**0.0031**	0.2123	0.986	0.014	0.0034	0.0113	0.2278	0.953	0.047	**0.0102**
PIN	0.0269	0.2564	0.905	0.095	**0.0251**	0.0339	0.3265	0.906	0.094	**0.0275**
Sentence	0.0350	0.1531	0.814	0.186	**0.0202**	0.0533	0.1867	0.778	0.222	**0.0443**

Table 4 presents the outcomes of the evaluation of Italian real and virtual users. For the real users the fusion only on signature leads to a small improvement of *1.2%*, while PIN and sentence lead to a decrease of approximately *4.5%* each. Regarding the virtual users' test data the fusion results are better than the single results of the modalities for all semantics. Here a relative improvement was reached for signature by *14.3%*, for PIN by *9.9%* and for sentence by *18.4%*. Another point of interest is the fact that the Indian users reach an EER of *0.0031* for the written signature. This is a value which is more than ten times better than the values of German users ($EER_{signature}=0.0350$) or Italian users ($EER_{signature}=0.0328$).

Table 4. Weights and EERs of Italian real and virtual users (H=handwriting subsystem, S=speech subsystem)

	Real users					Virtual users				
	Single system		Fusion			Single system		Fusion		
Semantic	**EER_H**	**EER_S**	**$weight_H$**	**$weigths_S$**	**EER**	**EER_H**	**EER_S**	**$weight_H$**	**$weigths_S$**	**EER**
Signature	0.0328	0.2933	0.900	0.100	**0.0324**	0.0529	0.2743	0.838	0.162	**0.0463**
PIN	**0.0380**	0.4429	0.921	0.079	0.0398	0.0333	0.3818	0.920	0.080	**0.0303**
Sentence	**0.0219**	0.3453	0.940	0.060	0.0229	0.0174	0.2971	0.945	0.055	**0.0147**

Finally, table 5 shows the results of the joint test sets of German, Indian and Italian donors. One can see that the fusion results of real users as well of virtual users are better than results of single systems in all cases. On the other side the results based on fusion of real users' data lead to smaller EERs than virtual data. Here the relative discrepancy for signature is *14.4%*, for PIN it is *36.2%* and for sentence it is *20.7%*.

Table 5. Weights and EERs of all real and virtual users (H=handwriting subsystem, S=speech subsystem)

	Real users					**Virtual users**				
	Single system		**Fusion**			**Single system**		**Fusion**		
Semantic	**EER$_H$**	**EER$_S$**	**weight$_H$**	**weigth$_S$**	**EER**	**EER$_H$**	**EER$_S$**	**weight$_H$**	**weigth$_S$**	**EER**
Signature	0.0102	0.2803	0.965	0.035	**0.0101**	0.0140	0.2950	0.955	0.045	**0.0118**
PIN	0.0321	0.3527	0.917	0.083	**0.0294**	0.0466	0.3474	0.882	0.118	**0.0461**
Sentence	0.0250	0.2406	0.906	0.094	**0.0219**	0.0325	0.2518	0.886	0.114	**0.0276**

All relative changes in respect to EER in the test sets mentioned above are shown in table 6. Improvements are denoted by a negative value and aggravations are denoted by a positive value. An example: If one has a look at the row *Signature* of German donors, one can see that the EER based on handwriting becomes better by *49.57%* by using virtual users (see column *H_{real} vs. $H_{virtual}$*). The next column shows that the single result based on the virtual users' speech data is better than result from real user data. Column *$Best_{real}$ vs. $Fusion_{real}$* presents the relative improvement of the best single subsystem (always handwriting) in comparison to the multimodal fusion result. In case of the signature of German donors the EERs of the best single system and the fusion yield an identical value of *0.0350*. This comparison in regard to virtual database is presented in column *$Best_{virtual}$ vs. $Fusion_{virtual}$*, where the fusion reaches a better result than the best single system with a relative improvement of *5.41%*. The last column *$Fusion_{real}$ vs. $Fusion_{virtual}$* shows the improvement of the fusion results of real and virtual users; here a relative improvement of *57.66%* was determined by using the virtual database.

Table 6. All improvements/aggravations grouped by real and virtual database, nationality and semantic

Semantic	**H$_{real}$ vs. H$_{virtual}$**	**S$_{real}$ vs. S$_{virtual}$**	**Best$_{real}$ vs. Fusion$_{real}$**	**Best$_{virtual}$ vs. Fusion$_{virtual}$**	**Fusion$_{real}$ vs. Fusion$_{virtual}$**
German Donors					
Signature	-49.57%	-9.61%	0.00%	-5.41%	-57.66%
PIN	10.25%	13.77%	-12.11%	-6.18%	15.00%
Sentence	-154.36%	21.31%	-15.66%	-8.56%	-138.74%
Indian Donors					
Signature	72.57%	6.80%	8.82%	-10.78%	66.67%
PIN	20.65%	21.47%	-7.17%	-23.27%	8.73%
Sentence	34.33%	18.00%	-73.27%	-20.32%	54.40%
Italian Donors					
Signature	38.00%	-6.93%	-1.23%	-14.25%	30.02%
PIN	-14.11%	-16.00%	4.52%	-9.90%	-31.35%
Sentence	-25.86%	-16.22%	4.37%	-18.37%	-55.78%
Joint Set of Donors					
Signature	27.14%	4.98%	-0.99%	-18.64%	14.41%
PIN	31.12%	-1.53%	-9.18%	-1.08%	36.23%
Sentence	23.08%	4.45%	-14.16%	-17.75%	20.65%

In the comparison of the results of the *Joint Set of Donors* (see table 6) one can see that the results determined for real users of the handwriting subsystem are always better than the results of virtual ones. The relative difference lies between 23% and 32% for the three semantics. At the speech subsystem only for the PIN a small relative improvement (1.53%) was reached by using virtual data; for the signature and sentence the real users based EER are better than the values achieved by virtual users. For the joint set the multimodal fusion has reached in both cases, real and virtual database, an improvement for all three semantics. The relative improvement based on real users' database is between 1% and 15% and for virtual users it is between 1% and 19%. However, the multimodal fusion results obtained by real users are better than the results determined from virtual users.

Based on the results presented in table 6 no general recommendations regarding creation and/or use of virtual users can be given. In the presented test environments an alteration of properties of the test set by extending the number of individuals and by changing the cross relations between the modalities of each user leads to unexpected results. It can not be predicted whether the creation of a virtual multimodal database, based on real multimodal and single-modal biometric data, will be result in an improvement or aggravation of the authentication performance.

5. CONCLUSIONS AND FUTURE WORK

This paper has suggested a new evaluative methodology to study the impact of using virtual multimodal database for multimodal biometric experiments. By this way we have presented three methods to create a virtual database. We have validated one out of the suggested three concepts by performing experiments for a bi-modal fusion system based on handwriting and speech. Unfortunately no regularities were found to predict the behavior of a virtual multimodal database containing single and multimodal biometric data. A general observation, in our test scenarios the fusion based on virtual users leads always to an improvement of the verification performance of the single subsystems. On the other side, the fusion of real users' modalities reached a better result in comparison to the single algorithms in only *8* out of *12* cases. However, in only 4 out of 12 cases the fusion based on virtual database results in a better performance than the real users' fusion.

One aim of our further work will be the collection of additional handwriting and speech data in order to enhance the test databases and obtain an equal number of users for the cultural based test sets. In future the behavior of virtual multimodal databases will be studied and compared with the results presented in this paper in order to research the influence of shuffling only real users' multimodal data to the authentication performance of the bi-modal system (see figure 4 (iii.)). Another point is to create virtual multimodal databases by usage of well known single and/or multimodal data, for example free available biometric data used for biometric evaluations.

ACKNOWLEDGMENTS

The work described in this paper has been supported in part by the European Commission through the IST Programme under Contract IST-2002-507634 BIOSECURE. This publication has been also produced partly with the assistance of the European Union, project CultureTech. The content of this publication is the sole responsibility of the University Magdeburg and their co-authors and can in no way be taken to reflect the views of the European Union.

We would particularly like to thank Franziska Wolf for her support and fruitful discussions in the context of our work.

REFERENCES

1. A.K. Jain, A. Ross, "Multibiometric Systems", Communications Of The ACM, Vol. 47, No. 1, 34-40 (2004).
2. C. Vielhauer, S. Schimke, A. Valsamakis, Y. Stylianou, "Fusion Strategies for Speech and Handwriting Modalities in HCI", Proceedings of SPIE-IS&T Electronic Imaging, Vol. 5684, 63-71 (2005).
3. C. Vielhauer; T. Scheidat, "Multimodal Biometrics for Voice and Handwriting", J. Dittmann, S. Katzenbeisser, A. Uhl (Eds.), Communications and Multimedia Security: 9th IFIP TC-6 TC-11International Conference, CMS 2005, Proceedings, LNCS 3677, 191-199 (2005).
4. T. Scheidat, C. Vielhauer, J. Dittmann, "Distance-Level Fusion Strategies for Online Signature Verification", Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Amsterdam, The Netherlands (2005).
5. F. Wolf, T. Scheidat, C. Vielhauer, "Study of Applicability of Virtual Users in Evaluating Multimodal Biometrics", Lecture Notes in Computer Science, Volume 4105/2006, Springer Berlin / Heidelberg, 554-561 (2006).
6. C. Vielhauer, T. Basu, J. Dittmann, P.K. Dutta, "Finding Metadata in Speech and Handwriting Biometrics", Proc. of SPIE-IS&T Electronic Imaging, Vol. 5681, 504-515 (2005).
7. A.K. Jain, S.C. Dass, K. Nandakumar, "Soft Biometric Traits for Personal Recognition Systems", Proceedings of International Conference on Biometric Authentication, LNCS 3072, pp. 731-738, Hong Kong (2004).
8. F. Wolf, T.K. Basu, P.K. Dutta, C. Vielhauer, A. Oermann, B. Yegnanarayana, "A Cross-Cultural Evaluation Framework for Behavioral Biometric User Authentication", 29th Annual Conference of the German Classification Society (2005).

9. T. Scheidat, F. Wolf, C. Vielhauer, “Analyzing Handwriting Biometrics in Metadata Context”, Security, Steganography and Watermarking of Multimedia Contents VIII, edited by Edward J. Delp III, Ping Wah Wong, Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6072 (2006).
10. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, http://www.cdt.org/privacy/eudirective/EU_Directive_.html
11. C. Vielhauer, *Biometric User Authentication for IT Security: From Fundamentals to Handwriting*, Springer, New York, 2006
12. C. Vielhauer, R. Steinmetz, A. Mayerhöfer, “Biometric Hash based on Statistical Features of Online Signature”, Proc. of the Intern. Conf. on Pattern Recognition (ICPR), Conference on Pattern Recognition (ICPR), Quebec City, Canada, Vol. 1, 123-126 (2002).
13. S. S. Stevens, J. Volkmann, E. B. Newman, “A Scale for the Measurement of the Psychological Magnitude Pitch”, Journal of the Acoustic Society of America, Vol. 8, 185-190 (1936).
14. J. W. Tukey, B. P. Bogert, and J. R. Healy, “The frequency analysis of time series for echoes: cepstrum, pseudo-autovariance, cross-cepstrum and saphe cracking”, Proceedings of the Symposium on Time Series Analysis, 209-243 (1963).
15. Y. Linde, A. Buzo, and R. Gray, “An algorithm for vector quantizer design”, IEEE Transactions on Communications, Vol. 28, 84-95 (1980).
16. The Culture Tech Project, Cultural Dimensions in digital Multimedia Security Technology, a project funded under the EU-India Economic Cross Cultural Program, http://amsl-smb.cs.uni-magdeburg.de/culturetech/